

Pulling rank

Why should US military personnel be singled out for genetic discrimination?

It might seem hard to believe, given the current contrast between the conditions endured by US military personnel in Iraq and elsewhere, and the comfort and extravagance of people back home, but the US Congress is close to passing a bill that will deprive men and women in the military of a right that will be enjoyed by everyone else. The law would make the military the only group of individuals who can be discriminated against on the basis of genetic tests for health conditions.

Although it is currently being held up by Senator Tom Coburn (Republican, Oklahoma), apparently in league with the health-insurance industry (see *Nature* 448, 631; 2007), the Genetic Information Nondiscrimination Act is widely expected to pass. When it does, it will establish in law an important principle that has long been sought by geneticists and ethicists: that people should not be at risk of losing their health insurance or their jobs if they take a genetic test.

In a nation where health-care benefits can be lost on the flimsiest of excuses, the law is essential if genetic tests for disease are to develop into useful tools (see *Nature* 448, 2; 2007). Extraordinarily, however, the Department of Defense has managed to exempt itself from the bill. This has been achieved without public debate and, indeed, almost by rote to please generals who are putting their own logistical convenience above the interests of the men and women under their command. As the *Los Angeles Times* reported earlier this month, the defence department already discriminates shockingly against soldiers whose health problems it can attribute to genetic disposition, sometimes discharging them without benefits.

There's plenty of precedent for military exemptions in legislation — the Pentagon customarily opts out of all kinds of laws and regulations. Politically, it is easy to do. Self-described 'supporters of the military' in Congress just wave the flag and dig their heels in, usually barking that the US military won't be constrained by some liberal wimp who wants to, say, impose fuel-emission standards on military vehicles.

But in this case, Congress seems to have taken its marching orders

from the senior brass, even when they conflict directly with the interests of men and women in the ranks. This is a disgrace. The argument for the exemption, such as it is, holds that the military has to beware in case its health-care provisions turn it into a haven for people who know they are ill and connive to milk these provisions once they have been signed up.

It is — to put it very, very mildly — improbable that someone seeking to malingering into a cosy health-care plan would chose, at this juncture, to sign up for service in the US military. They would be far more likely to go and work, for example, on Capitol Hill, where they could loaf around all day drafting preposterous clauses for insertion into otherwise sensible legislation. And where, when illness strikes, they could cash in on the relatively generous health-care benefits afforded to congressional staff.

The civilian branches of the government, like every other US employer apart from the Pentagon, will be barred under the legislation from terminating employment or health coverage on the basis of results from genetic tests. At the moment, most such tests provide imprecise and provisional indications of susceptibility to disease. They will grow more valuable in flagging appropriate interventions (such as surgery or drug treatment) when they can be used in appropriate circumstances without the threat of discrimination.

Even with the exemption in place, it would be straightforward for the defence secretary, Robert Gates, to address the matter immediately. Gates doesn't have to wrestle with the complexities that make it tough to control the behaviour of different actors in the wider society. All he has to do is issue a command that, henceforth, no member of the US military will be discharged or otherwise discriminated against on the basis of genetic testing. It's that simple. ■

"It is improbable that someone seeking to malingering into a cosy health-care plan would chose to sign up for service in the US military."

More than lip service

Biology and physics have much to offer each other — but they must forge equitable partnerships.

Towards the end of the twentieth century, many distinguished people hailed the arrival of the 'century of biology'. Physics was, in the minds of some, given a pat on the head and sent packing. But now, a hunger for funds, a fashion for multidisciplinary research and genuine intellectual interest are increasingly driving biologists into the welcoming arms of other disciplines.

Experiments on single biomolecules (such as those described on page 984) represent just such a trend. In the 1990s, alert experimen-

tal and theoretical physicists noticed that these experiments could provide them with intriguing polymers to explore in relatively cheap bench-top applications. Their involvement offered biologists a new way to scrutinize some of the leading actors in life's haphazard play. Predictive models have been developed on the basis of reproducible, quantifiable experiments. The interplay between life's codes and its physical constraints is being uncovered.

Applying first principles inside the cell remains an enormous challenge, although classic examples of cell physiology provide inspiration. Patch-clamp technology, for example, opened the electrophysiology of neurons to modelling and theory that will continue to chip away at the mysteries of signalling in the brain.

Now, significant multidisciplinary progress can be anticipated in understanding how physical forces shape the inner workings of cells.

For example, researchers recently examined glial cells, an enigmatic type of cell found in the brain, and by analysing the mechanical properties of the cells found that they could not act as glue or as support for neurons as had previously been believed (Y. B. Lu *et al. Proc. Natl Acad. Sci. USA* **103**, 17759–17764; 2006).

Other examples have included a predictive model for spindle alignment — a step that tells cells how to divide — based on physical forces inside the cell induced by its adhesion to a surface (M. Théry *et al. Nature* **447**, 493–496; 2007), and an analysis of cytoskeleton behaviour in response to cell stretching (X. Trepat *et al. Nature* **447**, 592–595; 2007).

In such endeavours, efforts must be made to ensure that a collaboration is truly intellectually productive for all disciplines involved. The initial urge may be for biologists to go to physicists or mathematicians for help in developing techniques or building models to answer purely biological questions, creating a one-way relationship. Alternatively, the allure of simple, elegant models may have some theorists working to ends that don't necessarily provide biological

insight. But in the best examples of interdisciplinary work, insight and enlightenment are mutual. Biologists get a chance to answer key questions in their field while mathematicians and physicists develop and apply tools that better inform their understanding of the natural world. Otherwise, calling such work 'interdisciplinary' is little more than lip service.

If cell biologists are truly to engage physicists and vice versa, a better sense that both are in this ride together is necessary. The papers mentioned above involve exploring physical forces acting on a cellular scale. Marrying those measurable physical forces to cellular chemistry in a meaningful way promises to push biology far beyond today's biochemistry. It is a challenge that could engage research for decades. And physics, in particular, is needed more than ever. ■

"Marrying measurable physical forces to cellular chemistry in a meaningful way promises to push biology far beyond today's biochemistry."

Space for capitalism

Rich people who play with rockets should be encouraged, but not subsidized.

Although British prime minister Edward Heath turned a fresh phrase in castigating "the unacceptable face of capitalism" in the 1970s, he was hardly unearthing something new. Aspects of capitalism have always suffered from unpalatable appearances, sometimes coinciding with genuine flaws.

But capitalists' knack for opening up markets and creating wealth has benefited society sufficiently to make some of its practitioners' faces more than acceptable. Few fit more squarely in that camp than those who have made their fortunes through computers and the Internet. "The largest single legal creation of wealth we've witnessed on the planet", as venture-capitalist John Doerr has termed it, was brought about by imaginatively finding ways to provide things that made lives and businesses more efficient, more effective, more fun, or some combination of all three.

Now a few of these people are devoting some of their acquired fortunes to the as-yet-untested business of inexpensive space flight (see page 988). Jeff Bezos, the founder of Amazon.com, has gathered together the expertise he thinks is needed to build rockets that will fly passengers first to the edge of space, later to orbit. Elon Musk, one of the begetters of PayPal, is building a range of rockets, some tailored to traditional satellite markets, some to taking people to the International Space Station. SpaceShipOne, which three years ago won the Ansari X prize for flying to an altitude of 100 kilometres and back twice within a fortnight, did so with the financial backing of Paul Allen, one of the founders of Microsoft. SpaceShipTwo is being developed in partnership with the Virgin Group, chaired by Richard Branson, a popular capitalist from a different background.

At the very least, this activity is likely to provide some thrills for wealthy customers — and cheaper launch options for certain types

of satellite. Today's established rocket companies are vast concerns deeply embedded in the military-industrial complexes of various nations; it is a fair assessment that entrepreneurial competition will shake them up a bit.

At best, one or more of these companies might actually find ways to make the launch of private citizens into orbit cheap and routine. This wouldn't just allow a lot of people to fulfil their childhood fantasies; it would also make it cheaper for governments to put people into orbit — a capability that a number of them currently maintain at very high cost for little clear benefit. And it would render the eventual exploration of other bodies in the Solar System more affordable than it is today.

This somewhat distant prospect, however, should not obscure various grounded truths. One is that getting cheap, reusable vehicles into orbit and back again is not going to be easy, and may well prove beyond the reach of current technologies.

There are also security concerns. Given that the technologies needed to circle round Earth are basically the same as those needed to lay waste to the ground below, their development cannot always be viewed as an unmitigated good. Some faces would be entirely unacceptable as owners of what amounts to a privatized, intercontinental ballistic missile. The issue of who decides what constitutes 'acceptable' in that context remains unresolved.

Finally, it can be anticipated that some would-be space entrepreneurs will, given half a chance, seek subsidy from the public purse. Such calls should be treated with scepticism. Certain public-private partnerships may make sense, and the programmes so far offered by NASA to encourage the development of private-sector resupply craft for the space station seem to do so. But in general, those who believe in private spaceflight should pursue their dream at their own expense. ■

"It is a fair assessment that entrepreneurial competition will shake established rocket companies up a bit."

RESEARCH HIGHLIGHTS

Venom not a digestif

J. Exp. Zool. doi:10.1002/jez.411 (2007)

The presence of digestive enzymes in snakes' venom has prompted scientists to propose that their venom evolved to help them consume their prey. Experiments carried out by Marshall McCue at the University of Arkansas in Fayetteville suggest that this is not the case.

McCue found that western diamondback rattlesnakes (*Crotalus atrox*, pictured) digested mice no faster when the mice had been injected with venom than when they had not. The diamondback's venom is among the richest in digestive enzymes, so this result lends support to theories that snakes evolved venom for some other purpose — such as hunting or defence.



J. VANDYKE

PALAEONTOLOGY

Climate for change

Geology **35**, 831–834 (2007)

Mineral deposits from caves in southern Israel (pictured below) point to increased rainfall in this now inhospitable region between 140,000 and 110,000 years ago, researchers report.

The timing coincides with the dates of early human sites in northern Israel. This supports the idea that climatic changes helped modern humans to migrate from Africa by making the deserts suitable for human settlement and passage.

There was already evidence for parts of the Saharan desert being wetter during this period, but the new data, from Anton Vaks at the Geological Survey of Israel and the Hebrew University of Jerusalem, Israel, and his colleagues, covers the 'bottleneck' region of the Sinai-Negev Desert, which connects

Africa to Asia. A subsequent return to arid conditions in this region may have spurred further northward migration.

GENETICS

Indicators of asthma

Am. J. Hum. Genet. doi:10.1086/521200 (2007)

Researchers in the United States have shown how a difference in one genetic base, or single nucleotide polymorphism (SNP), may influence a person's risk of developing asthma.

Several SNPs found in non-coding regions near the gene *HLA-G* have been associated with asthma risk. Zheng Tan at the University of Chicago, Illinois, and his team looked in these regions for docking sites for genetic regulators known as microRNAs. The group found one docking site that comes in two versions, differing in a SNP dubbed +3142C/G. One version binds microRNAs better than the other. Tan's group showed that this SNP, which is situated close to those previously associated with asthma, also influences asthma risk, probably through effects on microRNA binding. The effect depends on maternal asthma status, which suggests that differences in microRNA regulation begin in the womb.

CELL BIOLOGY

Transformers

Cell **130**, 678–690 (2007)

Centrioles are the 'Transformers' of the cell: in one guise, these organelles form structures known as centrosomes that help organize chromosomes during cell division; in another, they seed the formation of cilia and flagella.

Brian Dynlacht and his colleagues at the New York University School of Medicine have now identified two proteins that may

help centrioles switch between these roles.

Cells have only one pair of centrioles, so they cannot have cilia and divide at the same time. The team found that boosting cells' levels of a protein called CP110, known to have a role in the centrosome cycle, suppressed cilia formation. The researchers also showed that cells lacking either CP110 or Cep97, which they characterize as recruiting CP110 to centrosomes, grew extra cilia.

MICROBIOLOGY

A light touch

Science **317**, 1090–1093 (2007)

A light-sensing enzyme that helps plants turn to face the Sun has been found in four species of bacteria, including a human pathogen. The enzyme, a histidine kinase, is kicked into action by blue light.

Roberto Bogomolni, at the University of California, Santa Cruz, and his colleagues scanned bacterial genomes for similar enzymes and identified contenders in a marine bacterium, a plant pathogen and in two species of *Brucella* — a pathogen of cows and humans.

Brucella that lack the enzyme or that have been kept in the dark are more likely to be killed by immune cells than normal *Brucella*. The benefit to the bacterium of linking its virulence to light remains unclear.

BIOTECHNOLOGY

Seeing red

Nature Methods doi:10.1038/nmeth1083 (2007)

An engineered protein should offer biologists images of deep tissues in live animals better than any they can produce today.

Animal tissues are most transparent to light from the far-red end of the visible spectrum but, unfortunately for scientists



A. VAKS

seeking to light creatures up from the inside, no natural proteins fluoresce very well at these wavelengths. Dmitry Chudakov of the Shemiakin-Ovchinnikov Institute of Bioorganic Chemistry in Moscow, Russia, and his colleagues created and tested more than 100,000 mutant forms of a red sea anemone protein to find one with a longer-wavelength glow. This protein, dubbed Katushka, was brighter than two existing far-red fluorescent proteins in tests with transgenic frogs. The researchers have also made a monomeric version of the protein, mKate, to use as a tag on other proteins.

PHYSIOLOGY

An eye on the time

Cell **130**, 730–741 (2007)

Switching off the eye's circadian 'clock' disrupts visual processing in mice, researchers have found. Many organs of the body have such clocks, but their function has been unclear.

Charles Weitz at Harvard Medical School and his colleagues deleted a circadian control gene, *Bmal1*, in the retina of mice. They noted that this disturbed the normal night and day cycle of retinal electrical activity.

By contrast, disrupting the central circadian clock in the brains of wild-type mice did not significantly affect the rhythm of the retinal impulses. This suggests the retinal circadian clock has an autonomous role in visual processing. Extensive changes in light-dependent gene expression are thought to be associated with the effect.

OPTICS

Stuck in a loop

Phys. Rev. A **76**, 023816 (2007)

Light ricocheting around inside a silica sphere like sound in a whispering gallery can be brought to a stop, researchers have shown.

Anatoliy Savchenkov and his colleagues at the Jet Propulsion Laboratory in Pasadena, California, calculated that a series of pulses of light entering a 'whispering-gallery-mode' resonator — a device that accepts only certain wavelengths of light — can end up with a zero 'group velocity'. This theoretical prediction runs against general expectations, but the team confirmed it through experiments.

Researchers have proposed using other systems that slow light for information storage. Savchenkov says it should be possible to build a resonator memory, although the stopped light in the resonator as configured in these experiments could hold no information.

STEM CELLS

Hope for the broken-hearted

Nature Biotechnol. doi:10.1038/nbt1327 (2007)

Poets tell us that, once broken, the heart does not readily heal. But progress in repairing damaged rat hearts with human embryonic stem cells raises hopes that such treatments may one day work for humans.

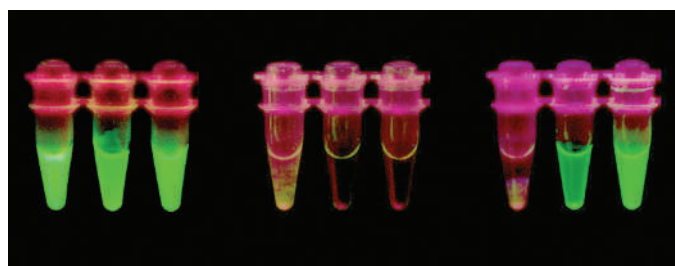
Charles Murry at the University of Washington in Seattle and his colleagues report that human embryonic stem cells treated with two proteins, activin A and bone morphogenic protein 4, develop into heart cells more efficiently than those treated with current methods. They also concocted a 'prosurvival cocktail' that blocked potential cell-death pathways, improving the survival of stem-cell derived heart cells when they were transplanted into rats. In rats that had had a heart attack, these cells replaced dead tissue, and prevented deterioration of heart function.

CHEMICAL BIOLOGY

Unscrambling the egg

J. Am. Chem. Soc. **129**, 10110–10112 (2007)

Anyone who has boiled an egg will have noticed the solidification that occurs as heat causes proteins to unfold then aggregate. Such aggregation can be a problem for the therapeutic use of proteins, so scientists are



interested in schemes that might prevent it.

David Liu and his colleagues at Harvard University, have found a way to make various proteins resistant to aggregation. They 'supercharge' the protein surfaces by substituting amino acids that appear on the outside surface of the folded protein with amino acids that can be ionized. Surprisingly, this does not seem to interfere with the protein's structure or function — but it keeps the molecules soluble after boiling, whereas unmutated forms aggregate.

The picture (above) shows solutions of the biochemical marker green fluorescent protein before (left) and immediately after boiling (centre), and after cooling (right). The leftmost vial in each set contains the standard protein, the other two contain supercharged forms.

AM. CHEM. SOC.

JOURNAL CLUB

Joe F. Costello
University of California, San Francisco, USA

To an epigeneticist, cancer is encrypted in genes and their packaging.

Early in my career I had the good fortune to study epigenetics in a lab focused on the molecular genetics of cancer. At the time, geneticists typically thought that in cancer, epigenetic changes — which affect regulation of the genome but not the genome's sequence — were epiphenomena less worthy of study.

This might have made the experience akin to being a Republican mayoral candidate in left-leaning San Francisco; instead it was positively transforming.

As my own research group took shape, I began to integrate genetic and epigenetic theories of malignant transformation. Now, hereditary human cancers and genetically engineered mice once held up as evidence for genetic models also provide evidence for epigenetic models, and we study the interactions of the two mechanisms.

In this light, a recent paper (G. G. Wang *et al. Nature Cell Biol.* **9**, 804–812; 2007) captured my attention because it dissects how one genetic change leads to epigenetic changes that ultimately cause leukaemia.

The work focuses on an abnormal fusion protein — produced after part of one gene fused, or translocated, with part of another — and narrows down its cancer-causing properties to a particular region of the protein. This region mediates an epigenetic change: it adds a methyl group to one amino acid of a histone, part of a gene's packaging in the nucleus.

The team found that the fusion protein misdirects its methylation to the histones that package *HoxA* genes, triggering further miscoding of the histones. This activates the genes, which promote self-renewal of blood-cell precursors, contributing to leukaemia.

I wonder if the interactions could be traced back even further. Given the role of epigenetics in stabilizing chromosomes; might it have been epigenetic miscoding that made the gene susceptible to translocation in the first place?

Discuss this paper at <http://blogs.nature.com/nature/journalclub>

NEWS

Air force had early warning of pulsars

It was one of the most important astronomical discoveries of the twentieth century, and it became one of the more controversial when only one of the discoverers received a Nobel prize. Now a fascinating new footnote has been added to the story of how pulsars were discovered with the revelation that some had previously been observed by a US Air Force staff sergeant at a remote Alaskan outpost.

Earlier this month, 81-year-old Charles Schisler came forward to tell the story of how he used a military radar to identify around a dozen radio sources, some of which were pulsars. Astronomers who have seen Schisler's meticulous logs believe that he spotted a bright pulsar in the nearby Crab Nebula months before the first scientific observation of a pulsar was published in *Nature* (A. Hewish *et al.* *Nature* 217, 709–713; 1968). Although Schisler never knew exactly what he was seeing, the story should be counted as an early pulsar spotting, says Jocelyn Bell Burnell, an astronomer at the University of Oxford, UK, and one of the authors on the original paper. "He happened to be a very observant person," Bell Burnell says.

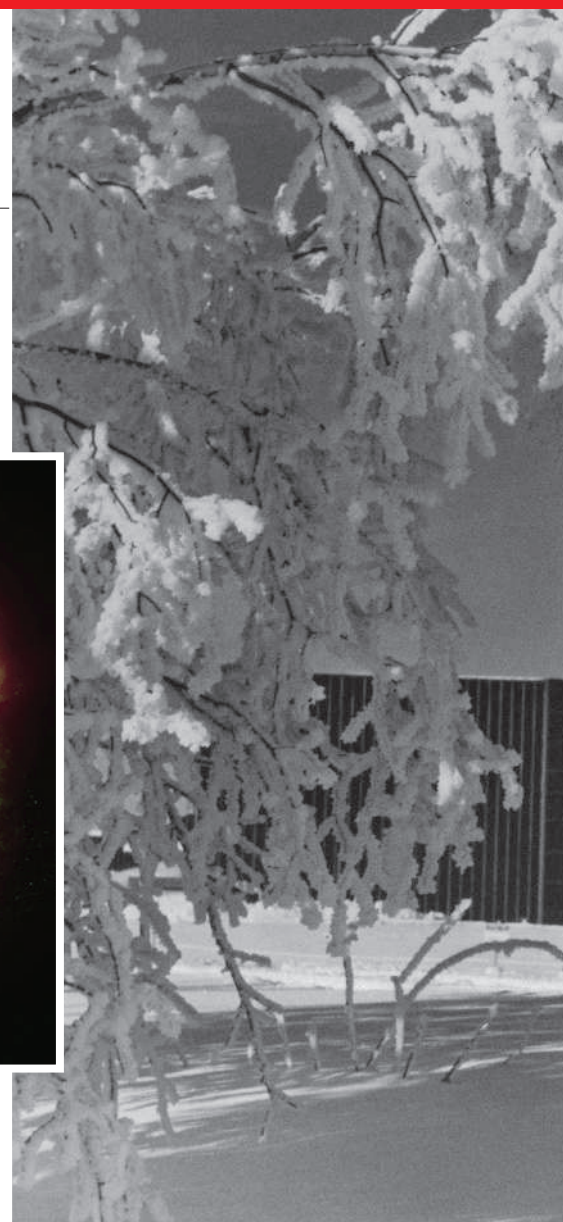
Schisler's story began in the summer of 1967 during a mind-numbing four-hour shift on an early-warning radar at Clear Air Force Station in Alaska. He was using the Ballistic Missile Early Warning System, a massive set of radars that looked some 4,800 km across Siberia for incoming warheads. As he sat at his station, the bored, 41-year-old staff sergeant noticed a faint signal on his scope. During the course of the



summer, the source continued to show up. "I kept seeing it week after week," he says.

Then one day Schisler noticed something — the mysterious blip appeared 4 minutes earlier than the day before. Four minutes meant a lot to the airman: before being stationed at Clear, he had been a navigator on a B-47 bomber, and he knew that stars rise 4 minutes earlier each night as a result of Earth's motion around the Sun.

Schisler calculated the source's approximate position in the sky and wrote it down on a scrap



of paper. Over the weekend, he drove 125 km to the University of Alaska at Fairbanks, where he met an astronomy professor who showed him the location of his source — the Crab Nebula, some 6,300 light years from Earth. At the centre

Market takes a gamble on carbon futures

The global carbon market is far busier this year than last, according to figures released this month. Around 1.2 gigatonnes of carbon — worth €15.8 billion (US\$21.4 billion) — were traded worldwide in the first half of 2007, primarily on the European Union (EU) Emissions Trading Scheme, according to analysts at carbon-market consultants Point Carbon in Oslo, Norway. Last year saw a mere

€22.5 billion change hands all year. But another Point Carbon analysis points out that the infant US market of ten northeastern states looks to be planning too many initial allowances.

Volume is up in the EU market, experts say, because traders are looking to the future. Trading on emissions allowances from the first, experimental, phase of the exchange collapsed in April 2006

and has never recovered. Most trading now is on allowances for the second phase, which begins on 1 January 2008 and runs to 2012. Currently, a credit bought on the open market for a tonne of emissions from the developing world through the Kyoto Protocol's Clean Development Mechanism costs less than buying the straightforward allowance to emit a tonne. As a result, many companies

in the industrial world are selling off the maximum permissible percentage of their allowances and making up the difference with these 'certified emission reductions' (CERs). This lowers prices (see *Nature* 448, 401; 2007) but increases volume. It has also led to a secondary market in buying and selling CERs. Another inducement to trade now is that companies can save allowances from the second



was designed to pick up man-made pulses bouncing off incoming missiles rather than steady signals, Schisler believes most of the things he saw were pulsars. By his own count, he spotted about a dozen sources. "My commanders didn't know what the hell I was doing," he recalls.

The work preceded by several months the observations made by Bell Burnell, then at the University of Cambridge, UK, which led to the first paper on the subject. A Nobel prize for the discovery was subsequently awarded to her supervisor Antony Hewish, but, controversially, not to her. Schisler was not the only one to "pre-discover" a pulsar, though, according to Bell Burnell. "There are actually a lot of stories," she says. In the 1950s, a woman visiting the observatory at the University of Chicago, Illinois, pointed out that there was a regularly pulsating source of visible light in the Crab Nebula. Elliot Moore, an astronomer at the university, dismissed the woman's claim, telling her that all stars seem to flicker. Another radio astronomer she knows of will, after a drink or two, confess to having dismissed observations of a pulsating source as the result of faulty equipment. "He's a bit embarrassed now," says Bell Burnell.

For his part, Schisler says he never quite understood what he was looking at until he heard of Hewish and Bell Burnell's discovery on a short-wave radio. When he learned that they had discovered a pulsating radio star, Schisler says, the significance of his own work became clear. But he says he didn't dare speak about the log until nearly half a century later, when the old early-warning system at Clear was finally decommissioned. He says that he feels he deserves no credit for his work, but he still regrets that he was unable to share what he had seen. "I wish we had had a way to communicate with the scientific community," he says. ■

Geoff Brumfiel

An early-warning radar in Alaska picked up pulses from the Crab Nebula (left) before astronomers did.

of the Crab, a supernova remnant, there sits a bright pulsar.

Schisler returned to Clear with the coordinates of other radio sources that he thought

he might be able to detect with the radar. Throughout the late summer and early autumn, he began a meticulous log of anything he could spot on the scope. Because the radar

phase for use in any year in the future, a feature called 'infinite banking'.

The increased activity is good for the health of the market, according to Guy Turner, director of New Carbon Finance in London. For one, it smooths out ups and downs in price, calming skittish investors. "It's the weight of numbers," says Turner. "Things move much more sluggishly." And, according to Endre Tvinnereim, senior analyst at Point Carbon, increased volume is good for the planet as well. "Increased activity increases the number of

actors with a financial stake in a better environment," he says.

Across the Atlantic, as the US Congress lurches towards passing a bill that would establish a carbon market in America, a few vanguard states have started their own. But analysts say that this scheme may face the same problem that crashed the European market. The value of phase I credits in the EU scheme fell to nearly nothing because there were far more credits floating around than companies needed. Point Carbon's analysis indicates that the US Regional Greenhouse

Gas Initiative (RGGI), comprising ten states on the northeast coast, from Maine to Delaware, may be planning allowances for 13% more carbon than they need. Meanwhile, Arizona, California, New Mexico, Oregon, Washington, Utah and two Canadian provinces have agreed to develop a cap-and-trade scheme within the next 12 months.

Janet Peace, an economist at the Pew Center on Global Climate Change in Washington DC, points out that the number of credits in the RGGI had been set before the EU phase I crash. "RGGI had already

put their model rule out there. I don't think you can blame them for the lesson that came later." She also says that over-allocation will not necessarily spell doom for the nascent US market, which is to start trading in 2009. Many states will auction off their credits, some with a minimum bid, thus giving them some value from the off.

But there is a lesson for future schemes in the over-allocation crash and subsequent rebound in the EU, Peace says: "You need good data on which to hang your cap." ■
Emma Marris

NUMBER CRUNCH

33% is the proportion of Americans following science and technology news "very closely" during 1986–89, according to the snappily titled Pew News Interest Index.

16% is the corresponding figure for 2000–06.

17% is the number of news junkies with a keen interest in celebrity scandals, showing that the world of science is now measurably less popular than Paris Hilton, though perhaps not as much as Sidelines might have thought.

ZOO NEWS

Bridge birds

A build-up of corrosive pigeon guano on the support struts of the bridge in Minneapolis that collapsed on 1 August might have played a role in the disaster. Some experts think that ammonia in the droppings could have weakened the steel beams.



J. INGLIS/ALAMY

WORDWATCH

Slime-Snake-Monkey-People

Evangelist Robert Bowie Johnson has coined this term for darwinists. He suggests that Christians should use it to 'shame' those who accept evolution over Genesis. Talk us through the snake-to-monkey step, would you, Bob?

SCORECARD



Chewing gum

Archaeologists at Harvard University have obtained 2,000-year-old DNA from wads of plant matter that Native Americans apparently used as chewing gum.



Chewing gum

Meanwhile, the oldest-ever 'chewing gum' has been found in Finland. The 5,000-year-old morsel, complete with toothmarks, was a distinctly unappetizing lump of birch bark tar.

Sources: Pew Research Center, Associated Press, Christian News Wire, Science, The Guardian

Cheaper approaches to flu divide researchers

As governments race to stockpile Tamiflu and other antivirals in preparation for the next influenza pandemic, scientists are starting to look at a different approach — drugs that modulate the immune system.

Proponents argue that these abundant and cheaper drugs, such as statins, might be a useful option where antivirals are unavailable or have been given too late. But sceptics fear that they might do more harm than good because they suppress the body's immune response to the virus. With the latest studies seeming to contradict each other, the flu community is divided about the best way to proceed.

The high mortality associated with viruses such as avian flu is thought to be because the immune system has a massive over-reaction to the infection. The body hikes its production of immune molecules called cytokines, which flood into the bloodstream, causing what's known as a cytokine storm. This triggers inflammation and lung damage, and can lead to multiple organ failure and death. So rather than the usual approach of targeting the virus, researchers hope that mortality could be reduced by using drugs that regulate this immune response.

At present, planning for pandemics is focused on stockpiling antivirals. But because these drugs, such as Tamiflu, are expensive and difficult to manufacture, there may not be enough to go round by the time another pandemic strikes, and poorer countries may not have access to them at all.

"If we are to be serious about confronting a pandemic, we must urgently evaluate agents that are already available and affordable to people in all countries," says David Fedson, former director of medical affairs at Paris-based Aventis Pasteur, now Sanofi-Aventis.

This month, Ian Clark's group at the Australian National University in Canberra published a study¹ investigating the effectiveness of another immune-modulating drug, called gemfibrozil, in mice infected with the influenza virus. Gemfibrozil is a fibrate, a class of drug widely prescribed to lower lipid concentrations

in the blood. But it also inhibits the release of inflammatory cytokines such as tumour-necrosis factor, interferon- γ and interleukin-6, and stimulates production of the anti-inflammatory cytokine interleukin-4. Clark's team infected mice with H2N2 — the flu strain that caused the 1957 pandemic. Four days later, once the rodents had become sick, they injected 46 mice with 60 milligrams per kilogram of body weight of gemfibrozil once a day for six days. Twice as many of the treated mice (52%) survived compared with controls (26%).

Tread with caution

"The paper should spark some interest, but even more caution," says Erik De Clercq, a virologist at the Catholic University of Leuven in Belgium. He points out that it is difficult to

extrapolate from Clark's H2N2 mouse model to the H5N1 strain in humans. For example, he says, it did not show that increased survival was directly due to inhibition of the cytokine storm. Clark agrees and is planning further studies to rule out the possibility of confounding effects, such as the drug having a direct antiviral effect.

However, two other studies suggest that the cytokine storm might not be as fatal as thought. In one study², mice that had one cytokine pathway knocked out showed no reduction in mortality compared with normal mice after they were infected with H5N1. "These results refute the

popular paradigm that the cytokine storm is the cause of death during H5N1 infection," says Robert Webster, who carried out the study with his colleagues at St Jude Children's Research Hospital in Memphis, Tennessee.

In the second study³, by Jacqueline Katz's group at the Centers for Disease Control and Prevention in Atlanta, Georgia, mice deficient in the inflammatory cytokine interleukin-1 receptor had worse mortality, higher viral loads and more inflammation than did controls when infected with the HK/486 virus.

But both studies tested only three to five mice in each experiment, and looked at only a few of the many overlapping cytokine path-



"We must urgently evaluate agents that are already available and affordable to people in all countries."

REUTERS/CORBIS



THE GREAT BEYOND

Check out our new blog, rounding up science stories from around the world.

<http://blogs.nature.com/news/thegreatbeyond/>

<http://blogs.nature.com/news/thegreatbeyond/>



Viral overload: the 1957 flu pandemic affected millions of people around the world.

ways, points out Eric Mortensen of the Audie L. Murphy Veterans hospital in San Antonio, Texas. They need to be replicated with a much larger number of mice, and examining other potential pathways, says Mortensen.

Clark even asserts that “if the same outcomes had been repeated at $n = 15$, the statistics would show the opposite of what Webster’s group claimed,” he says. However, Rachelle Salomon, the lead author of the study, says “large numbers of mice would be ideal, but due to our experience with these viruses, I am confident with the result and conclusions.” The groups were small because the experiments needed to be done in laboratories with an approved biosafety level of at least 3 (the maximum is 4), she explains. Because Australia has different standards, Clark was able to use level 2 facilities.

Statins — another lipid-lowering class of drug with anti-inflammatory effects — might be a better bet than gemfibrozil, says Mortensen, who has papers in the press showing that these drugs lower the number of deaths from pneumonia. And then there is a preliminary analysis⁴ of 20,000 patients older than 50 years enrolled in the University Medical Center Utrecht primary-care network in the Netherlands by Eelko Hak and his colleagues at the university. They found that during flu epidemics, people who took statins were 28% less likely to develop res-

piratory disease and 51% less likely to die from any cause than were those who did not take the drugs. Another, case-control study⁵ of 130,000 patients also found that current use of statins lowered mortality from pneumonia.

But Frederick Hayden of the World Health Organization’s global influenza programme says that pandemic flu involves protracted viral replication, often over weeks, so using immune drugs when viral replication is high could do more damage than good. A more sensible idea might be to study joint treatments with immune modulators and antivirals, he says.

But definitive answers can only come from clinical trials done immediately after the onset of a pandemic, says Peter Openshaw of Imperial College London. “We need to be able to plan ahead and set up trial protocols, but that’s not easy to do when we don’t know when an epidemic may strike.”

Declan Butler

1. Budd, A. *et al. Antimicrob. Agents Chemother.* **51**, 2965–2968 (2007).
2. Salomon, R., Hoffmann, E. & Webster, R. G. *Proc. Natl Acad. Sci. USA* **104**, 12479–12481 (2007).
3. Szretter, K. J. *et al. J. Virol.* **81**, 2736–2744 (2007).
4. Hak, E., Verheij, T., van Essen, G., Bonten, M. J. M. & Hoes, A. 16th European Congress of Clinical Microbiology and Infectious Diseases, Nice, abstr.; www.blackwellpublishing.com/eccmid16/abstract.asp?id=49073 (2006).
5. Schlienger, R. G., Fedson, D. S., Jick, S. S., Jick, H. & Meier, C. R. *Pharmacotherapy* **27**, 325–332 (2007).

Flying insects threaten to deafen Japan

A cicada known as the kumazemi is descending on Japan en masse, deafening the citizens and wreaking havoc on the country’s fibre-optic system. The 6- to 7-centimetre-long black cicada (*Cryptotympana facialis*) inhabits western Japan and subtropical regions of eastern Asia. This August is expected to be the largest and noisiest cicada summer in the insect’s four-year cycle.

Surveys of the shells shed by maturing cicadas at a park in Osaka, carried out by Shigehiko Shiyake, curator of the Osaka Museum of Natural History, and Eiji Numata, a biologist at Osaka University, show that the cicada population increases every year for four years, after which it returns to base level and the cycle restarts. From the past three years’ data, the scientists calculate that this year will be the four-year peak, with nearly 2.5 times as many cicadas as in 2006. The noise level is also set to climb. Measured at 90.4 decibels at another Osaka park last year, this year the same spot is expected to hit 94 decibels — decibels follow a logarithmic scale, so that’s more than double the volume. Prolonged exposure to this level of noise can cause deafness.

The kumazemi are also cutting households off from their Internet. Apparently mistaking fibre-optic cables for withered branches, they have been punching their one-millimetre-diameter ovipositors into the cables and laying eggs. In at least 1,000 cases over the past two years, the cicadas have either severed the cable or opened up a hole, allowing water to seep in. The Osaka-based Nippon Telephone and Telegraph West Corporation has responded by creating new cables that lack the grooves that the cicadas target with their ovipositors and by adding another protective plastic layer to the cable.

David Cyranoski



H. NUMATA

Q&A

Space, science and the bottom line

In April, planetary scientist **Alan Stern** joined NASA as associate administrator for science, putting him in charge of the agency's \$5.5-billion science budget. Now he must juggle more than 90 space missions and 3,000 grants with focuses ranging from Earth to the distant Universe.

You've announced extensions to a number of missions lately. Does this mean you're cutting future missions?

Not at all — it's an effective way to get more science done with our budget. Often this gives you more bang for the buck by providing important new science for very little additional cost. With these extensions, we're going to get two new comet visits and a search for extrasolar planets for about 15% of the cost that either of the Stardust and Deep Impact missions took to build. And there will be more future missions.

What about costs?

Cost can be held down by properly scoping missions to begin with and then closely managing to that budget. There's been too much management by open chequebook, and that will not be continued. We are changing the culture about overruns and we are not going to tolerate them as was done in the past. The Kepler mission [to find Earth-like planets] is one example. When they came saying they needed more money I said 'no' repeatedly. And they found a way to stay within budget.

The Mars programme is the biggest planetary programme at NASA, but what's the status of a mission to bring back rocks?

Mars sample return has been on the to-do list for 35 years, but it's had a very frustrating life cycle. It has repeatedly become very popular and then grown to be so large as to be unaffordable. But because it's been so highly ranked by the scientific community, I want to see us finally do it. So I've asked Doug McCuistion, our Mars programme director, to draw up architectures for an affordable Mars sample-return mission, and to design it to cost. To remain affordable it will have to be a Chevy sample-return instead of a Cadillac version, but I know what we've got from overly ambitious, Cadillac-like Mars sample-return mission plans: zero grams of martian rock.

Will it really happen this time?

To show people I'm serious about getting started, I've asked the spacecraft builders at the Jet Propulsion Laboratory to add a sample-caching ability to the our next rover, the Mars Science Laboratory [planned to launch in 2009]. As that drives around it will pick up samples and store them for possible pick-up by the sample-return mission a decade hence. This will give us critical practice at sampling on Mars, and it will also give us the option of skipping a future rover to cache samples — meaning we can reduce the cost and shorten the time until a documented sample is back on Earth.

Some scientists question the Vision for Space Exploration, NASA's push to return astronauts to the Moon. What real science is there to be learned from such a programme?

Let's remember that Apollo was very important scientifically. It taught us that the Earth-Moon system was born in a giant impact, it taught us about the importance of magma oceans, and its samples may even have recorded the signature of the clearing of the Kuiper belt [a phase in the early Solar System when mass was lost from the asteroid belt]. With twenty-first-century technology we will be able to accomplish a lot more science when we return to the Moon than Apollo ever could with 1960s technology. The vision for space exploration is critically important for the country, and it's also extremely exciting for science.

Imagine: we'll be sending sophisticated

field laboratories and, later, geologists, geophysicists and geochemists to the Moon. Perhaps the Vision will allow these same capabilities to be extended to near-Earth asteroids and to Mars. Think about Mars sample return with human crews — it will be tremendous. Instead of just hundreds of grams, humans can bring back hundreds of kilograms of samples. The science is just mouthwatering.

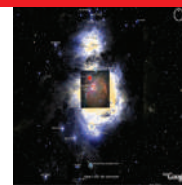
What about Earth? Will you restore some of the remote-sensing missions that have been cut in recent years?

We at NASA are currently operating a fleet of Earth-monitoring satellites that are the envy of other nations. We are simultaneously building the next Landsat mission, and we have seven other Earth-studies missions, representing billions of dollars in investment, under construction. No other nation can match this. We are now planning the next wave of Earth research missions recommended by the Earth-studies decadal survey released this year.

Interview by Alexandra Witze



P. E. ALERS/NASA


GOOGLE SKY PUTS ON A GREAT SHOW

Software tool great for the public; not perfect for amateur astronomers.

www.nature.com/news

NASA, ESA, DIGITIZED SKY SURVEY CONS., STSCI-GOOGLE PARTNERSHIP

No confidence vote for agency head

David Schwartz and the agency he heads are not having much of a summer. The National Institutes of Health (NIH) has announced that it will investigate “the management and leadership” of its environmental agency, the National Institute of Environmental Health Sciences (NIEHS) in Research Triangle Park, North Carolina. NIEHS director Schwartz is temporarily stepping aside “to eliminate even the perception that the review is anything other than independent and unbiased”. The investigation is expected to take several months.

Schwartz was a controversial figure even before he took the job. In April 2005, after the NIH announced new rules on the kinds and amounts of stock its employees could hold, Schwartz said he would not take the post of director. After confidential chats with NIH head Elias Zerhouni, he felt reassured enough to take the job.

Since then, a cavalcade of charges has been made against Schwartz, many of which came

from the office of Senator Charles Grassley (Republican, Iowa). Grassley accused Schwartz of overspending by \$4 million on his own lab at the expense of other intramural researchers and took exception to the expert-witness role Grassley took in some asbestos-related lawsuits, allowing attorneys to dwell on his credentials. Schwartz has since stopped giving expert testimony. Scientists at the institute also complain that he has disregarded the results of peer review when allocating funds and provoked conflict by attempting to privatize one of the institute’s journals, *Environmental Health Perspectives*.

And this month, officials at NIEHS handed out a form requesting that staff report any phone calls they receive from Congress — the type of form usually used by staff who deal with legislative affairs. On 20 August, Grassley wrote a letter to Zerhouni that said: “Handing this form out to rank and file NIEHS employees during the

course of a congressional investigation could cause these employees to feel that management is attempting to flush out whistleblowers or any other individual assisting me with my inquiry.” However, it is not clear whether the forms went to more than a handful of people.

Scientists in the institute’s intramural research programme have also voted on their feelings about the tumult. Of 206 tenured and tenure-track scientists, 146 participated. Of those, 99 said that the actions and decisions of Schwartz had affected their morale negatively; 91 said that Schwartz did not have their support; and 107 said that they had no confidence in him.

“Morale is at an all time low,” says a senior scientist at the NIEHS, who did not want to be named for fear of retaliation. “A lot of people grew up with this institute and it’s like their home, and he’s come and soiled its reputation.” ■

Emma Marris

“Morale is at an all time low.”

Boston's biodefence lab passes safety review

Plans for a biodefence research lab at Boston University do not pose an undue health risk, according to a review commissioned by the National Institutes of Health.

The study, conducted by the State University of New York at Buffalo, used models to simulate the accidental release of infectious agents from the lab. It found that Rift Valley fever, Ebola, Sabia and monkeypox would not spread any more rapidly in the populous South End district of Boston, where the lab is being built, than they would at other, less populous sites.

But David Ozonoff, an epidemiologist at Boston University who opposes the biosafety level-4 lab, questions the relevance of the simulations. "The agents they modelled aren't very contagious person-to-person," he says. The report's findings will come into play in ongoing state and federal lawsuits.

America to help fund circumcision in Africa

The US President's Emergency Program for AIDS Relief (PEPFAR) is to fund programmes to circumcise men as a way to

stem the spread of AIDS in Africa.

The decision is a reversal of the \$15-billion programme's previous stance, which did not allow spending for male circumcision. But officials say they have been persuaded by recent evidence showing that the technique can reduce the chance of men contracting HIV by as much as 60% in some African communities.

Still, some activists caution against hasty rollout of such programmes unless they are accompanied by additional studies and public health education (see *Nature* 447, 1040–1041; 2007). PEPFAR says that it will fund circumcision programmes only if they include counselling and HIV-testing services. Programme budgets are likely to start out at less than US\$1 million per country.

Turkish creationist secures block against blog site

A Turkish court last week blocked access to WordPress.com, a website that hosts more than a million blogs, many of which are science related. The order came after prominent Muslim creationist Adnan Oktar accused some blogs hosted by the site of containing defamatory speech.

Oktar, who writes under the name Harun

Yahya, previously made headlines for sending copies of his anti-evolution book *The Atlas of Creation* to scientists around the world.

Oktar apparently took action to block WordPress after comments attributed to Edip Yuksel, an opponent of his, appeared on the site. The case is the latest in a series of high-profile freedom-of-speech incidents in Turkey, including the criminal prosecutions of a number of journalists and authors.

Nobel laureate comes out of retirement for Salk

The Salk Institute for Biological Studies in La Jolla, California, has appointed Nobel



P. APRAMIAN/CORBIS

The Salk Institute: still searching for a president.

laureate Roger Guillemin as its interim president, after an almost year-long search for a chief executive proved unsuccessful.

The previous president, Richard Murphy, announced his resignation last November, and left on 1 July. Earlier this month, Murphy became interim president at the California Institute for Regenerative Medicine in San Francisco (see *Nature* 448, 739; 2007).

Guillemin, a Salk researcher between 1970 and 1989, is 83 and will come out of retirement to take over from 1 October, officials say, as the search for a permanent head continues.

The French endocrinologist won the Nobel Prize in Physiology or Medicine in 1977 for studies of brain hormones.

Sugar rush at the National Institutes of Health

The National Cancer Institute (NCI), part of the US National Institutes of Health (NIH), is betting that carbohydrates will provide useful cancer biomarkers. On 21 August, the NCI announced that it will spend \$15.5 million over the next five years on seven 'tumour glycome laboratories'.

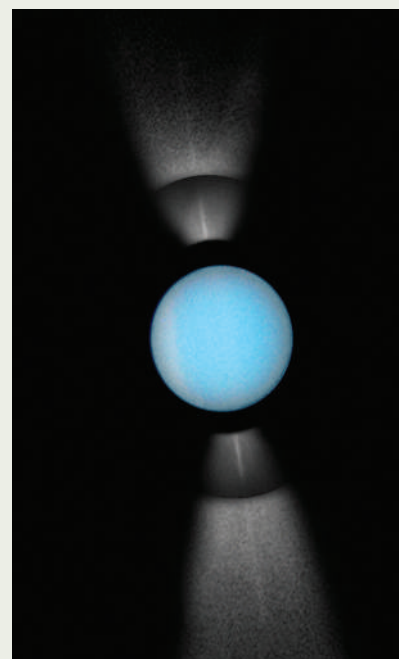
The labs are looking for sugar biomarkers to aid in the study, detection and early

Uranus offers rare glimpse of its ring system

NASA's Hubble telescope recorded a rare view of the complete ring system of Uranus earlier this month. Every 42 years, the rings line up edge-on to Earth, which allows astronomers a good view of the dust particles that make up the rings. Images from NASA's spacecraft Voyager 2 captured Uranus' rings in 1986, but since then the inner ring, zeta, which is usually concealed by others, has migrated closer to the planet, this recent sighting shows.

Such changes in structure have been observed for other ringed planets, but Uranus seems to be in a particularly dramatic cycle of change, possibly caused by meteorite impacts that kick up dust.

This image from the Hubble Space Telescope shows the rings as two spikes above and below the planet. Most of the glare has been blocked out, but a small amount remains as a fan-shaped artefact. Uranus, taken using a shorter exposure, has been superimposed on the rings.



NASA/ESA/M. SHOWALTER (GETTY)

diagnosis of breast, ovary, lung, pancreas, colon and blood cancers.

The awards cap a recent NIH spending spree on glycobiology through its trans-NIH Alliance of Glycobiologists for Detection of Cancer and Cancer Risk programme. Two other consortia in this alliance have already

been granted more than \$50 million from institutes within the NIH. One is focused on functional glycomics (and is partnered with Nature Publishing Group in the Functional Glycomics Gateway) and the other is developing a technology and resource repository.

BUSINESS

How to make it in Moscow

Life for scientific entrepreneurs in Russia can be fraught with difficulty. **Alison Abbott** meets one who has ridden the rapids for more than a decade.

It was 1989 when Vadim Govorun was sent to Berlin to learn more about genetics. He arrived from Moscow shortly before the fall of the Berlin Wall — an event that triggered the upheavals that shaped his career as one of Russia's more successful scientific entrepreneurs.

Today, Govorun is head of a 200-strong company developing and selling medical diagnostics kits. But he also has an active academic career and sometimes, he says, he feels he's being stretched a bit too far.

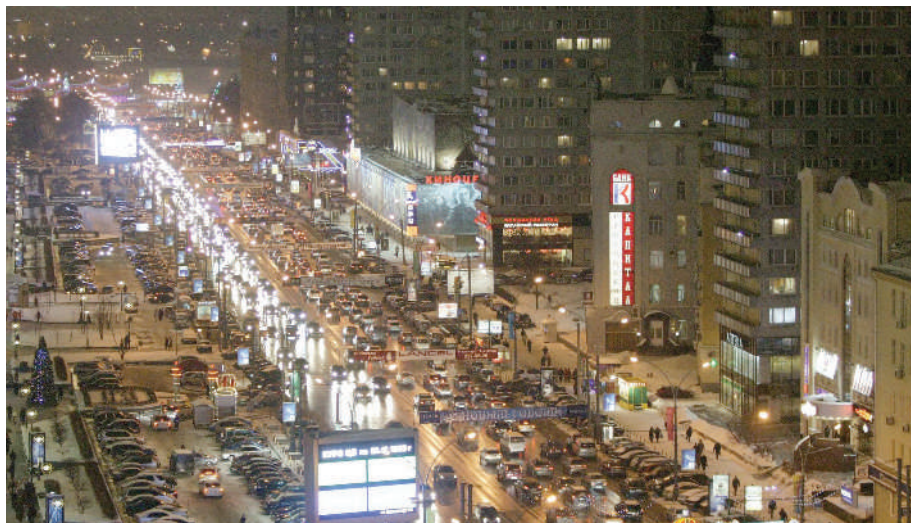
By 1994, when Govorun moved back to Russia from the Max Delbrück Centre in Berlin, the effects of the collapse of the Soviet system were clear to see. "When I came back I had no idea how to survive as a researcher," he recalls. He returned to the Institute of Physico-Chemical Medicine in Moscow, a branch of the health ministry, where salaries had been devalued to poverty levels — if they were paid at all. The institution could not even pay its electricity bill, never mind buy research materials.

Govorun spent the 600 deutschmarks (then US\$360) he'd scraped together in Germany on a Russian-made DNA synthesizer and set up shop. He made oligonucleotides, short pieces of DNA that are essential ingredients in diagnostic kits that identify disease-causing microbes from small samples of their genetic material. The institute let him use one of its rooms to develop the kits: it needed its researchers to build up income any way they could.

Researchers at the Centre for Obstetrics, Gynaecology and Perinatology at the Russian Academy of Medical Sciences in Moscow, provided him with pathogens that cause sexually transmitted diseases (STDs), and helped him to test the kits on samples from patients.

With plenty of demand from customers who couldn't afford to import diagnostics, Govorun's company, Lytech, grew steadily. It now sells to more than 150 hospitals and private STD-testing labs, a large number of which sprang up in the 1990s. "It became fashionable in Russia to test for STDs, but not for other important diseases such as cardiovascular disorders — I don't know why," says Govorun.

He profited from the trend, and has pumped



M. FOMICHEV/ITAR-TASS

Mean streets: corruption and burgeoning costs make Moscow a tough city for biotech entrepreneurs.

\$600,000 in revenue back into the firm to finance its expansion. Lytech added antibody-based ELISA diagnostic kits to its portfolio, and extended the range of pathogens for which it can test. It also makes 'biochips' to test for mutations, such as those that make bacteria resistant to antibiotics.



"When I came back to Moscow I had no idea how to survive as a researcher."

— Vadim Govorun

Lytech's story is typical of small biotechnology companies in post-Soviet Russia, which tend to be located at government institutes and to pay for their own expansion gradually, out of revenue. Its main Russian competitor in clinical diagnostics is InterLabService, founded a couple of years before Lytech by German Shipulin, a biologist at the health ministry's Centre for Epidemiology in Moscow. Still headquartered at the centre, it employs 600 people and has a factory in Moscow that makes the kits.

"Most of the successful biotechnology companies in

Russia started in this way, and financed themselves by reinvesting," says Shipulin, noting that there's no tradition in Russia of banks and other investors lending money to entrepreneurs.

Some start-ups have got help from Russia's new tycoons. Biochemist Vladimir Skulachev at Moscow State University set up Mitotechnology

to find anti-ageing drugs, with undisclosed millions of dollars of support from Oleg Deripaska, an aluminium magnate who is said to be Russia's second-richest man.

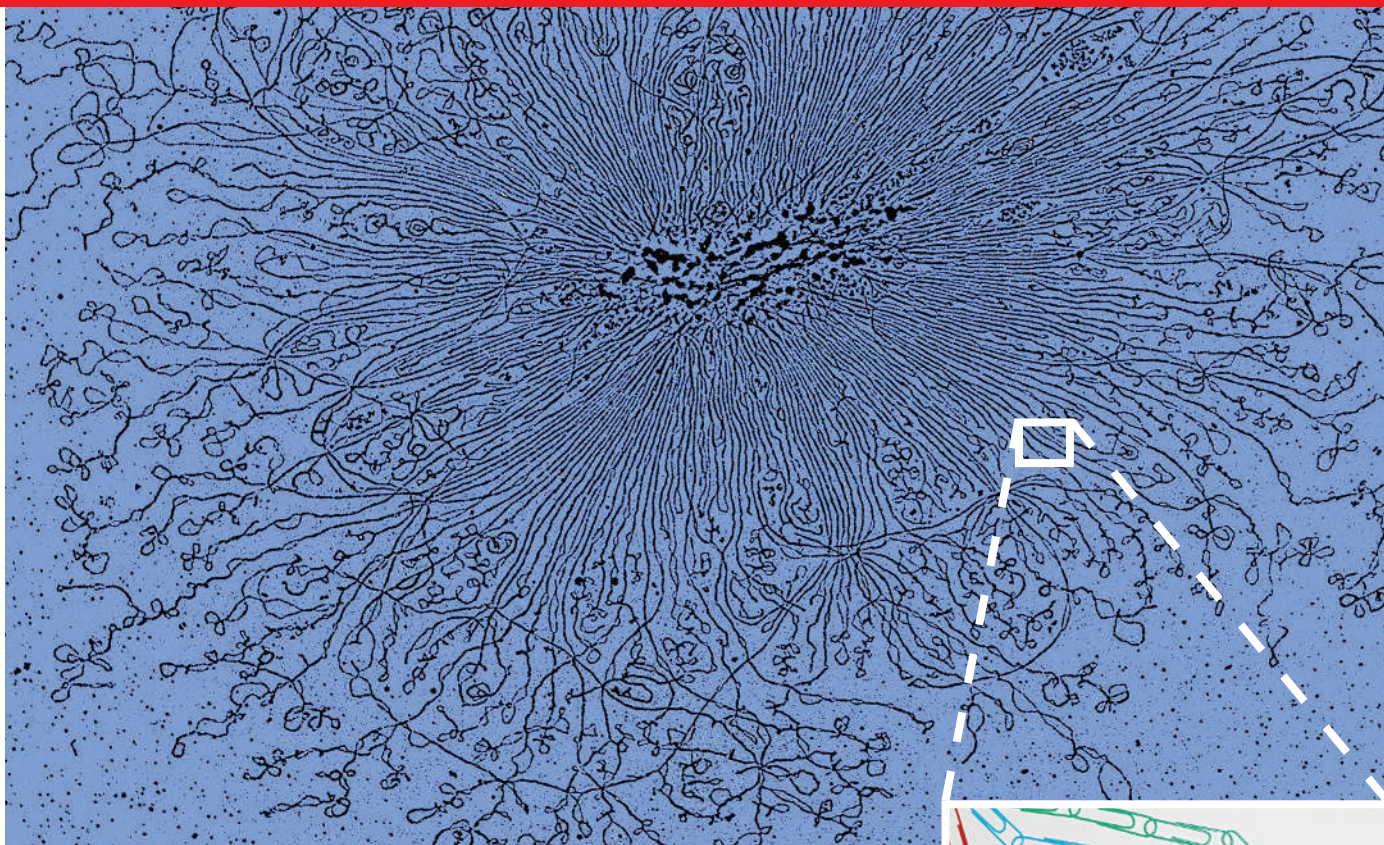
Torsten Wöllert, who deals with energy and environment at the European Union's office in Moscow, says the current crop of biotech firms emerged at a time when home-grown high-quality scientific labour was cheap, and impoverished research institutes were offering space to start-ups for relatively low rents. None of these conditions exists any more, he points out.

And the prevalent corruption has made businesses even more difficult to run, Wöllert says. The largest single asset a company can have is political cover, or *krisha* (literally 'roof'). Entrepreneurs "need contacts high up in the bureaucracy to prevent those lower down from making difficulties or asking for bribes", he says.

Govorun says he is careful to follow the ever-changing bureaucratic rules in order to avoid trouble. But after 12 years, his relationship with Lytech is at a cross-roads. Govorun is now deputy director of the Institute for Physico-Chemical Medicine and runs a large genetics lab there. He is also a professor at the neighbouring Physico-Technical Institute.

So it's decision time: he will either expand Lytech, perhaps opening facilities in cheaper cities than Moscow, and building up export sales, or he'll return to science full-time. "I haven't quite decided which way to turn — I'll be thinking about it all summer."

A. ABBOTT



COPYRIGHT DESIGNER GENES POSTERS LTD. DR. RUTH KAVENOFF 1944-1999

SPRING THEORY

Physicists interested in the mechanics of single molecules are helping open one of the blackest boxes in biology. **Brendan Maher** discovers how the disciplines are working together.

When trying to explain why he has become fascinated by physics lately, Kerry Bloom, a cell and molecular biologist at the University of North Carolina, Chapel Hill, pulls a handful of paper clips from his pocket and links them together. Stretching a small chain across the surface of his palm, he says: "Imagine this is DNA. You can stretch it to its full length, but each link in this chain is vibrating all the time." Bloom jiggles his hand, causing the paper clips to dance. The links twist at random, but once a couple of kinks or bends are introduced some force is needed to stretch the chain out again. After a few seconds of simulated brownian motion, the paper-clip chain collapses back into his hand.

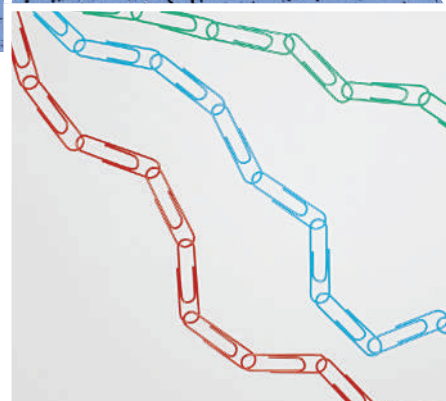
Bloom's paper clips are a demonstration of the properties of an 'entropic spring', a system where thermodynamics favours a resting state in which all the chain's components are bunched and tangled. Rubber bands and silk share these properties, and so does DNA. It is not a new insight, but to Bloom, who has spent most of his career speaking the language of genetics, it's a powerful one. He gets a wistful look when talking about it — he's even written poetry about it ("Francis and Jim lasted 50 long years/Isn't it time for some big new ideas..."). Spring theory,

as he calls it, might help explain a phenomenon he is deeply interested in — how the tiny biochemical machinery of the cell can manage billions of bits of information stored on vast polymer strings that need to be read, copied and packaged into an incredibly small space.

In the cartoon models that illustrate textbooks on cell- and molecular biology, purposeful proteins orchestrate neat, stepwise molecular dances as they react to coloured blobs and bind a perfect DNA staircase. Everyone finds their partner easily and does their job efficiently in a scale-free rendition of an otherwise empty space. The reality is something much more chaotic.

Eye of the storm

In the cell there's no eye-soothing white space to separate things. Water molecules are a constant omnidirectional hailstorm, van der Waals forces glue things together and viscosity rules. Within this molecular maelstrom, gravity is imperceptible, and there's more or less no inertia; all purposeful movement degenerates into random jittering the moment no further power is available. Bloom has a dramatic illustration of the strangeness: if a bacterium stops beating its flagella to move forward, it comes to a stop in "less than the width of a hydrogen bond", just a fraction of a nanometre.



CORBIS

Chains of life: DNA from a burst bacterium reveals a surprise at higher magnification.

For those who find these complications fascinating, the tools of modern physics are making them ever more amenable to study. Theoreticians and experimentalists are devising predictive mathematical models for the mechanical properties of cells at a molecular level, and starting to expose the formulae under which these tiny chaotic environments function. DNA, operating at the centre of this maelstrom, is of particular interest. "DNA is mechanically manhandled inside the cell," says John Marko, a condensed-matter physicist now working in Northwestern University's molecular-biology department in Evanston, Illinois — and that manhandling is important for replication, transcription, regulation, packaging and pretty much everything else DNA does or has done to it.

Oliver Rando, a biologist at the University of Massachusetts Medical School in Worcester who studies DNA packaging, hopes that physics may answer questions other approaches haven't touched. "You have these machines that appear all over the nucleus but only happen to

act at a couple of loci," he says, referring to specific places in the genome. "In some cases the detailed mechanism underlying that difference might be biophysical in nature." He's not certain that the physicists can solve the problem — but he's happy to see them try.

To find out how DNA works in the strange world of the cell, the first step is to look at how it behaves in simpler places. Carlos Bustamante, a pioneer in single-molecule biophysics at the University of California, Berkeley, got started in the field simply by thinking about the most everyday lab procedure: gel electrophoresis. DNA fragments loaded into a gel and then subjected to an electric field will migrate along the field lines, and the speed at which the different fragments do so reveals their size. In the late 1980s, when he was at the University of New Mexico, Albuquerque, Bustamante began to wonder about the details of the process, and used a microscope to watch fluorescently labelled DNA fragments migrating through a gel¹. "What was amazing was how elastic they were," he says. As the electrical field pulled on the negatively charged strands, they folded and curled, crawling like caterpillars through the gel's molecular obstacle course of crosslinked polymers.

Let's twist again

Bustamante started devising new experiments to stretch or twist the DNA and see how much force was needed to make the familiar double-helix structure break, unwind, or knot up like an old telephone cord. Key to these investigations were new and constantly improving ways of seeing and manipulating the molecular structures — by attaching beads of polystyrene to the ends or sides of long DNA molecules he could hold them in a magnetic field, or trap them with laser light and watch as the DNA squirmed and recoiled in reaction to what was done to it.

Because DNA is double-stranded and twisted, says Bustamante, it's quite rigid. But it also bends and folds — in fact it does so to an astonishing degree. The DNA in human cells is

packed so tightly that two metres of it squeezes into every nucleus. The trade-off between rigidity and flexibility depends on scale: on small scales the molecule seems stiff, on larger ones bendy. Jonathan Widom, at Northwestern University, compares DNA to a garden hose; easy to wrap around your waist, impossible to wrap around your finger.

The key to the difference between stiff and flexible is the chain's 'persistence length' — the distance that, as Widom puts it, "defines how far you need to go along a polymer before it forgets which way it was going". For a strand of DNA left to itself, studies have pinned the persistence length at about 50 nanometres, which corresponds to 150 bases; below this length DNA is difficult to bend. Results from dozens of studies fit fairly well a pre-existing 'worm-like chain' model of DNA, which predicts that it behaves somewhat like a chain of tiny paperclips.

But key cellular processes, including packaging and genetic regulation, require looping on a scale much smaller than 50 nanometres. And some experiments by Widom and his collaborators showed tiny sequences spontaneously forming loops — 'cyclizing' — at a much higher frequency than would be predicted by the worm-like chain².

These are the sorts of mismatch between theory and reality that excite Philip Nelson, a theoretical physicist at the University of Pennsylvania, Philadelphia. Nelson says he took notice of the work by people such as Bustamante and others in the mid-1990s because it put biological problems into a language he could understand. "If you knock out a gene and suddenly a rabbit doesn't like broccoli," he says, speaking as a physicist, "that's not helpful to us." DNA wrapping itself up in knots that the models seemed to preclude, though, was a problem he could get his teeth into.

A group including Nelson and Widom recently approached the problem of tight looping using atomic-force microscopy, a powerful visualization technique that allows them to look at the shape of DNA strands directly, rather than looking at

beads or fluorescence associated with them. They found that at lengths of between 5 and 10 nanometres (just 15 to 30 of the nucleotide subunits from which the double helix is built) the flexibility of the DNA was several orders of magnitude higher than that predicted by the worm-like-chain model. They proposed a new model called the sub-elastic chain³. Others, such as Marko and Jie Yan at the University of Illinois at Chicago had also been proposing models that allow for breakdown of the worm-like chain at

short scales⁴. But it's still a contentious area in the field. "You have to say that this is very much in flux at the moment," says Widom, who notes that some of the assumptions in his cyclization experiments have come into question⁵.

Widom has nonetheless found that he can make useful predic-

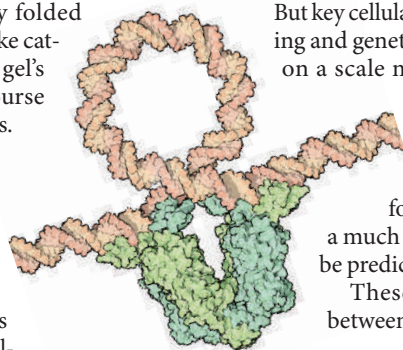
tions of the bending and looping proclivities of a piece of DNA on the basis of its sequence. "That makes a link between bioinformatics and mechanics," he says. A DNA section with a specific nucleotide pattern might bend more or less than another strand, and some proteins seem to read this 'code' from DNA's bendability rather than directly from the sequence. Histones, the barrel-shaped packaging proteins around which DNA winds in tight curls are a prime example. According to Widom, the histones seem to prefer specific DNA sequences based on their flex. Rando, who works on histone dynamics, says this is where physics influences his work: "That's a case where something super-important to biologists is directed at least partially by something biophysics-y."

And histones aren't the only proteins known to manage DNA looping. The lac repressor, a tiny V-shaped protein that grabs two specific sequences of DNA about 10 nucleotides apart and pulls them together, forms a very tight coil in the intervening material — too tight for any proteins that might want to unspool and read the DNA to cope with. Jeff Gelles, a self-described "dye in the wool *in vitro* biochemist" at Brandeis University in Waltham, Massachusetts, helped to develop a physical way to visualize this looping.

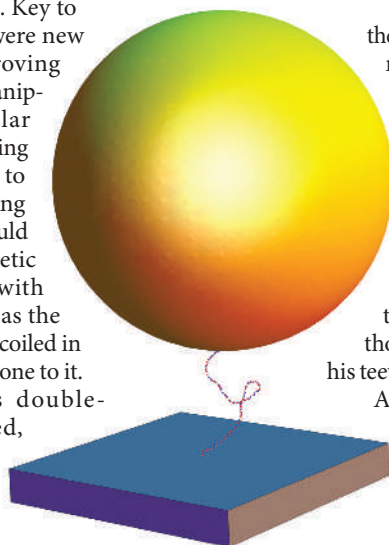
In 1991, he and his co-workers devised a simple way of looking at DNA mechanics visually⁶. They took a DNA strand that was being transcribed into RNA by a protein polymerase, fixed the polymerase to a glass slide and then attached a 40 nanometre gold bead to the free end of the DNA. The scale of the particle means that brownian motion has it dancing wildly — think balloon on a string in a hurricane — but the bead is big enough to see under a microscope. With time-lapse microscopy, the

"If you knock out a gene and suddenly a rabbit doesn't like broccoli, that's not helpful to us."

— Philip Nelson



The tight loop formed with the help of the lac repressor (green)...



...shortens the length of the DNA attached to this golden bead.

D. GOODSELL

K. TOWLES

researchers could extrapolate the length and movement of the DNA from the bead's random positions around a central tether point.

If these tethered particle experiments can reveal DNA length, they can reveal DNA looping, as the looping, by taking a hitch out of the tether, shortens its effective length. In unpublished work, Oi Kwan Wong, a former graduate student with Gelles, now at Stanford University in Palo Alto, California, used tethered particles to investigate how much the looping caused by a lac repressor shortened a sequence of DNA that was equipped with the relevant binding sites. The only problem is that she saw three different lengths: one stretched out, unlooped length and two looped lengths. "We think that the most likely explanation for that is not something to do with the structure of the DNA, but rather that the repressor itself can undergo a large change in the three-dimensional structure," says Gelliss.

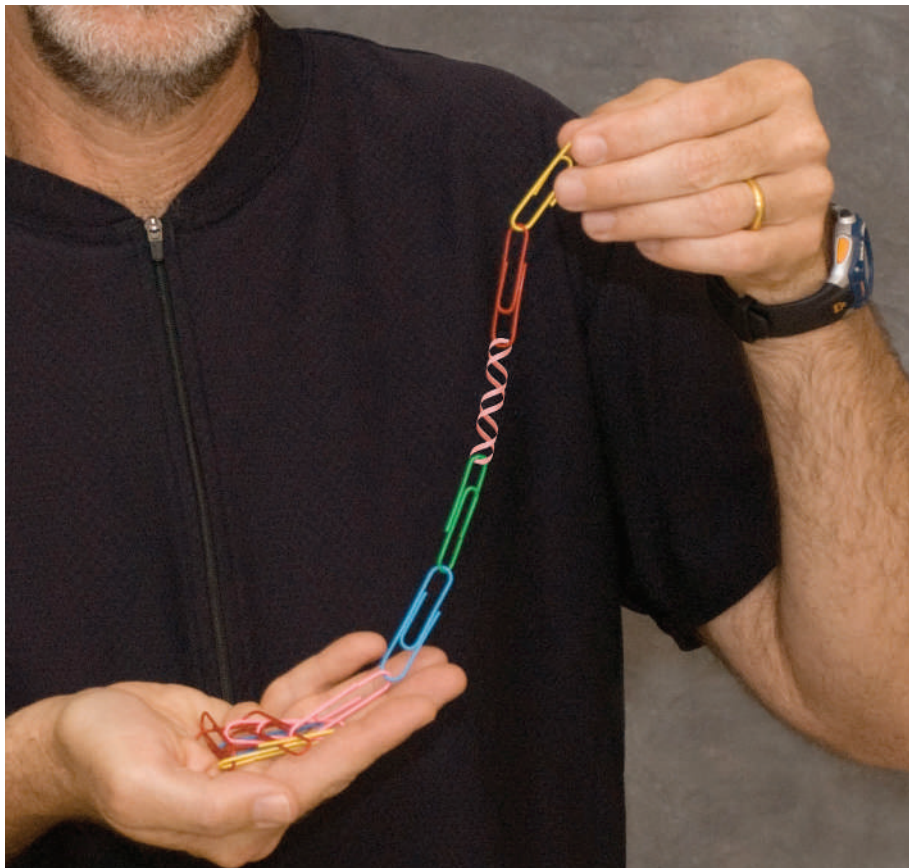
Internal workings

Rather than using microscopic gold beads, Bloom is trying to study the dynamics of DNA inside the cell itself — using the cell's own machinery to do the work. During cell division a protein-motor complex called a spindle separates identical copies of each chromosome, pulling one towards one end of the cell and the other in the opposite direction, allowing the eventual daughter cells to each get a complete set. The spindle latches on to the chromosomes at structures called centromeres, which have been the focus of Bloom's work for decades.

Some years ago he engineered a cell chromosome with a second centromere that he could turn on and off. When the extra centromeres are activated during cell division, the spindle will sometimes latch on to two centromeres on the same copy of a chromosome, stretching it across the cell rather than separating it from its twin. Occasionally, the stretched chromosome snaps like a wishbone, with the DNA recoiling to one end of the dividing cell.

Bloom originally used the method to study how the cell responds to DNA breakage⁷. But he is now starting to look at the dynamics of the break itself, using a laser to snip the stretched chromosomes and measuring the rate at which they recoil. "This is now where it gets complicated, and I'm not an expert," says Bloom. "We can see it stretch, we can see it recoil. How do I deduce force?" Elements of the worm-like-chain model predict his observations fairly well, he says, and with additional genetic manipulation, he can begin to look at how histones disassemble and reassemble in the stretched and recoiling DNA.

The moxie required to study DNA physics inside the cell itself elicits both admiration and scepticism. "I would say that it is really



B. NALLEY/S. WHITEFIELD

Resting state: paper clips and a touch of photoshopped DNA make their point in the palm of Kerry Bloom.

important to understand how these models work inside the cell," says Bustamante. "But let's not forget that the best biophysics is always done outside the cell." Hermann Gaub, a professor at Ludwig-Maximilians University in Munich, Germany, says he thinks observing systems that include some of the DNA's biological setting, but that stay away from the messiness of the cell proper are likely to be the most fruitful: approaches such as that of Marko, who pokes, prods and pulls the peculiarly large chromosomes of newts. "Doing it right inside the cell is what I would call heroic," says Gaub. But he's nonetheless intrigued.

Bloom admits that the work is preliminary. He says he's been boning up on polymer theory and is probing for potential collaborators — but he wants the data to show them first. "This is the stuff physics brings to the table," he says. The packaging of all that DNA into the nucleus of cell, he continues, "is another one of these big mysteries". A physical sense of how a hierarchy of folding patterns can pack DNA into the cell but also allow its sequences to be accessed remains far off. "By most estimations, it's packaged 10,000 fold," says Bloom, and no one knows how. "That's the attraction."

It's far from the only problem that physicists

have their eyes on. "People are now able to do extremely quantitative and extremely reproducible and precise experiments on individual biomolecules," says Marko, "and that's very attractive to physics people." And there are many biomolecules to choose from. To biologists, a text such as *Molecular Biology of the Cell* by Bruce Alberts and his colleagues is a well-trodden rendition of that which is known. But to a physicist first approaching biomolecules, it's an Aladdin's cave of shiny and captivating phenomena. Nelson says he's drawn to the sensation when something in biology "makes you ask 'How the heck does that happen?'" Physicists start reading Alberts' and they say that every three pages.

Brendan Maher is a features editor at Nature.

1. Gurrieri, S., Rizzarelli, E., Beach, D. & Bustamante, C. *Biochemistry* **29**, 3396–3401 (1990).
2. Cloutier, T. E. & Widom, J. *Mol. Cell* **14**, 355–362 (2004).
3. Wiggins, P. A. *Nature Nanotechnol.* **1**, 137–141 (2006).
4. Yan, J., Kawamura, R. & Marko, J. F. *Phys. Rev. E* **71**, 061905 (2005).
5. Du, Q., Smith, C., Shiffeldrim, N., Vologodskaya, M. & Vologodskii, A. *Proc. Natl Acad. Sci. USA* **102**, 5397–5402 (2005).
6. Schafer, D. A., Gelles, J., Sheetz, M. P. & Landick, R. *Nature* **352**, 444–448 (1991).
7. Thrower, D. A. & Bloom, K. *Mol. Biol. Cell* **12**, 2800–2812 (2001).

See Editorial, page 969.

DREAMS OF THE NEW SPACE RACE

When it was trying to catapult satellites into orbit the private launch business didn't get very far. Can it do better now that it's focused on giving the rich the ride of their life, asks **David Chandler**.

In the 1990s, the 'new space' vision was simple. Constellations of communications satellites in low-Earth orbit would require launch systems that could pop up little replacement satellites easily and cheaply. Small entrepreneurial companies would be able to meet the demand better than conventional launch providers such as Lockheed Martin, Boeing, Arianespace or even the cheaper options from the post-Soviet states. Unfortunately, the vision didn't get very far. The would-be contenders were short of capital, and the technological challenges were greater than appreciated. Worst of all, the vast expected market did not materialize.

In the past couple of years, however, the new-space sector has been booming, thanks to a change in its business model. Private spaceflight has always been a business for dreamers, and now one group of dreamers — the people building the spacecraft — has realized that another, much larger group of dreamers might be willing to pay a lot to ride on those spacecraft even if they can't do useful things such as get into orbit or launch satellites.

A company called Space Adventures, based in Vienna, Virginia, pioneered the marketplace for space tourism. Since 2001, it has sent five private individuals into orbit on Russian

spacecraft, at US\$20 million a trip. Then, in 2004, Scaled Composites of Mojave, California, became the first venture to build and launch a privately funded piloted rocket up to the edge of space. Entrepreneur Richard Branson immediately signed on. Starting in 2009, Virgin Galactic, a US subsidiary of the Virgin Group founded by Branson, plans to offer sub-orbital trips with a three-minute ride in weightlessness and an out-of-this-world view of black skies and curved horizons. At 1% of the fee for flying to the International Space Station, the trips are cheap only in a relative sense — but the number of people with \$200,000 to spend on such an adventure is growing, and the company says that it has taken at least \$25 million in advance bookings.

Dream teams

Numerous companies are now working on similar projects — a big change from the days of satellite dreams (see 'Private spaceflight: the different approaches'). "It's a very different situation now," says former space-shuttle astronaut Jeffrey Hoffman, now a professor

of astronautics at Massachusetts Institute of Technology in Cambridge. Even traditional industry giants, such as Paris-Based Astrium and Northrop Grumman in Los Angeles, California, are joining the game.

But these developments do not mean that the business will be an easy one, or risk free.

On the morning of 26 July, a loud blast ripped through the Mojave desert in southern California as a routine propellant test at Scaled Composites triggered an explosion. Three people died and three others were injured. But the industry as a whole seems not to have been badly shaken by the accident.

For one thing it was not a problem with the spacecraft that Scaled Composites is developing for Virgin — SpaceShipTwo — nor with its engines, but with gas storage. For another, the companies are all working with very different technologies. And for a third, most of the leading companies have relatively few investors to chase away — they tend to be dependent on the interest and investment of a particular multimillionaire or billionaire. As long as a small number of investors keeps the faith, the boom can continue.

"You really have to be able to take a lot of punches before you make it through the door."

— Robert Zubrin

ADRIAN DENNIS/AFP/GETTY IMAGES

Scaled Composites has a clear lead over the rest of the competition. No other company has yet managed to emulate the feat with which its SpaceShipOne, a small rocketplane launched from beneath a larger carrier jet, won the \$10 million Ansari X-prize. Scaled Composites is “the main chance here” for succeeding with a passenger-carrying rocket, says Robert Zubrin, an aerospace engineer from Colorado who in the 1990s worked to develop spaceplanes to serve the satellite market. Actually getting a vehicle into space, Zubrin says, “is a very tough thing. You put all this money in, and then something fails for the most trivial reason. You really have to be able to take a lot of punches before you make it through the door.”

Leader of the pack

Scaled Composites has already taken most of those punches — cushioned in no small part by the bankroll of Paul Allen, the co-founder of Microsoft, who paid for most of the \$20-million-plus cost of SpaceShipOne. SpaceShipTwo is meant to extend the success into a business with a small fleet of rocketplanes that can each carry two pilots and six paying customers. Whether that sort of business plan can sustain itself over the long term remains to be seen; the industry, however, is banking on it. Jane Reifert, president of the adventure-travel business Incredible Adventures in Sarasota, Florida, arranges trips for people who want to fly Russian MiG jets or cage-dive among great white sharks. She says that she has a long list of customers willing to pay for suborbital space flights. “People just want to do it,” she says. “They don’t care whether they’re the first or the three-thousand-and-first.”

Incredible Adventures has a contract to sell space rides with Rocketplane Kistler in Oklahoma City, Oklahoma — a company that is happy to make money from suborbital flights, but actually has its eye on getting people all the way to orbit. In this case, though, the people are government employees, not wealthy thrill-seekers. NASA’s space shuttle is set to be retired in 2010, and its replacement — the Ares/Orion system — won’t be ready to carry cargo and astronauts to the International Space Station until at least 2015. For those five years, NASA will have to rely on the Russians for transporting people and the Russians, Europeans and Japanese for transporting cargo. It would like to be able to supplement that capability with launches by private companies.

So the agency has set up a Commercial Orbital Transportation Services (COTS) programme. In the first phase of this programme NASA is providing up to \$500 million in the form of matched funds to private investments in two companies — Rocketplane Kistler and



The Falcon 1 rocket has yet to reach orbit.

Space Exploration Technologies (SpaceX) of El Segundo, California. The idea is to develop systems that can take supplies and, further down the line, crew members to the space station. In a major departure from the way NASA has always done business, the agency is specifying the end result but leaving design choices and related trade-offs to the companies; the companies will end up owning the technology and charging NASA, and others, for their services.

Of the two, SpaceX seems to be the most likely to make it. “Most knowledgeable people think SpaceX has a reasonable chance of success,” says John Logsdon, an expert in space policy at George Washington University in Washington DC. SpaceX was founded in 2002 by Elon Musk, who by the age of 30 had made hundreds of millions of dollars through Internet start ups, most notably PayPal. SpaceX has developed a small, single-engine rocket called Falcon 1, which has so far been launched twice: the first launch, in 2006, careened into the ocean; the second, earlier this year, soared to an altitude of 200 kilometres — well into space — but didn’t make it into orbit. The problem was traced to fuel sloshing in a second-stage tank. But all the most difficult parts of the mission — rocket ignition, liftoff, guidance, second-stage ignition and separation — worked without a hitch.

With the sloshing fixed, SpaceX plans to place a satellite into orbit with its third test flight in early 2008. And by the end of 2008, Musk hopes to launch a nine-engine, 54-metre-

high version that will ultimately be capable of carrying a crew of seven to the space station; on return, the crew capsule would parachute down to the ocean.

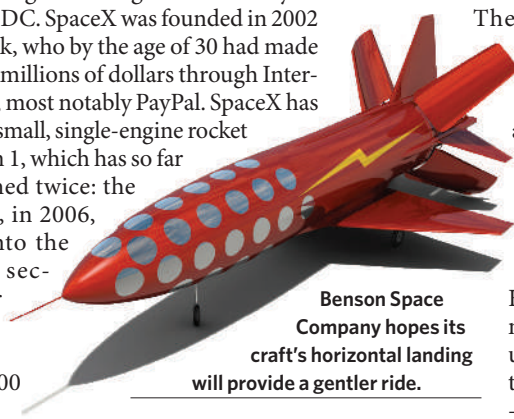
Musk hopes to get into orbit for about a tenth of the \$200 million cost of launching the similar-sized Atlas V, which is sold by a joint venture between Lockheed and Boeing. A key to making things cheaper, says Musk, is optimizing efficiency in every possible way. The Falcon craft has been designed from scratch, and every component — the propulsion, ground operations, electronic systems and overhead — has been streamlined as much as possible. “We’ve made substantial improvements that would be considered breakthroughs in all those areas,” says Musk. “There’s no one thing that is really the fundamental improvement.” To achieve all the cost savings he thinks possible, though, Musk will also need to find ways to reuse the rocket engines by bringing the spent booster stages safely back to Earth.

Off target

Rocketplane Kistler, the other potential COTS beneficiary, also hopes to make things affordable by making its stages fully reusable. Every part, including the crew capsule, is designed to parachute down to dry land for recovery. The struggling company could be at the end of its rope financially, however. It has missed several target dates for securing private financing, and the *Wall Street Journal* reported in late August that the firm had not been able to line up the institutional investors it needed to continue operations.

The company was formed in 2006 through the merger of Kistler Aerospace, which had been developing a two-stage orbital system similar to SpaceX’s, and Rocketplane, which had been working on a suborbital rocketplane similar to Scaled Composites’. Its K-1 design, a traditional multi-stage rocket not that unlike the SpaceX design, is the work of a space veteran — George Mueller, who was

head of NASA’s Apollo programme in the 1960s and lead designer of the space shuttle in the 1970s. K-1 also uses veteran rocket motors, albeit in a new configuration — Russian NK-33 and NK-43 engines, developed in the 1960s for the Soviet Union’s planned heavy-lift N-1 Moon rocket. The company’s chairman and chief executive officer, George French, disputes reports of its financial woes, telling *Nature*: “We will continue to pour in



Benson Space Company hopes its craft’s horizontal landing will provide a gentler ride.

BENSON SPACE COMPANY

SPACEX

PRIVATE SPACEFLIGHT: THE DIFFERENT APPROACHES

| Company | Project | Launch type | Fuel/oxidizer | Seats | Notable for |
|-----------------------------------|--------------------|---------------------------------------|-------------------------------------|-------|---|
| Armadillo Aerospace | Pixel (prototype) | Vertical take-off and landing | Ethanol/liquid oxygen | 2 | Entry in NASA Lunar Lander challenge |
| Blue Origin | New Shepard | Vertical take-off and landing | High-test peroxide | 4 | Preliminary test flights |
| Benson Space Company | X-1 | Vertical take-off, horizontal landing | Hybrid (rubber/nitrous oxide) | 6 | Designed to produce lower g-forces than other systems |
| Interorbital Systems | Neptune | Vertical take-off, straight to orbit | Unspecified hypergolic hydrocarbons | 6 | Fuel mix does not require ignition system |
| Rocketplane Kistler | K-1 | Vertical take-off, straight to orbit | Kerosene/liquid oxygen | 5 | Fully reusable orbital system |
| | XP | Horizontal take-off and landing | Kerosene/liquid oxygen | 4 | Jet engines for take-off, then rocket to reach space |
| SpaceX | Falcon/Dragon | Vertical take-off, straight to orbit | Kerosene/liquid oxygen | 7 | Test flight in March highest ever for private craft |
| Virgin Galactic/Scaled Composites | SpaceShipTwo | Horizontal take-off and landing | Hybrid (rubber/nitrous oxide) | 8 | SpaceShipTwo's predecessor, SpaceShipOne, was first civilian spacecraft |
| EADS Astrium | Astrium spaceplane | Horizontal take-off and landing | Methane/liquid oxygen | 5 | Jet engines for take-off, then rocket to reach space |
| Xcor Aerospace | Xerus | Horizontal take-off and landing | Kerosene/liquid oxygen | 2 | Rocket motors for take-off and to reach space |

our blood, sweat and money until the funding is secured.”

Many more companies are working on suborbital approaches. Although most are interested in rocketplanes, two are using an approach familiar to watchers of science-fiction movies but never before used for operational spacecraft: the same engine system for a vertical take-off and a vertical landing.

Straight up and down

Perhaps the most ambitious of these designs comes from Jeff Bezos, founder of Amazon.com and owner of a space company called Blue Origin, based in Kent, Washington. With a personal fortune estimated at more than \$3 billion, Bezos has the resources to develop a serious vehicle, and his company's motto — *gradatim ferociter*, “step by step with spirit” — suggests that he is in for the long haul. So does his choice of a vertical take-off, vertical landing design. Such an approach works best in vehicles with orbital and near-orbital performance, when the ability to slow down with rockets rather than with heat shields could be very handy.

Some clues about Blue Origin's near-term plans can be gleaned from an environmental

impact statement it had to file last year about its plans to launch from Bezos' ranch in west Texas. The company's first cone-shaped craft will be 15 metres high, automatically piloted and carry up to three passengers to heights of more than 100 kilometres. Passengers would experience about three minutes of weightlessness, and return to a landing pad just a few hundred metres from the take-off point. The company has been doing flight tests since November 2006.

Operating on a much less lavish scale, out of a warehouse near Dallas, Texas, Armadillo Aerospace was founded in 2000 by videogame designer John Carmack. Last October, the company — staffed in part by volunteers — fell just short of winning a prize NASA set up to inspire designs the agency hopes to use on the Moon, where vertical take-offs and landings are the only possibilities. The lunar landing challenge promises \$1 million to the first craft that can soar 50 metres high, hover for three minutes, land 100 metres away, and repeat the process in reverse within two and a half hours. Armadillo's small test rocket, named Pixel, has since met the requirements, but not yet under the scrutiny of judges. And, in the second ‘new space’ accident of the year, Armadillo's other vehicle, Texel, crashed and

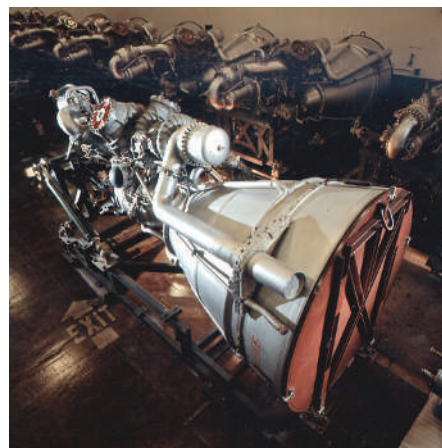
burned during a test flight on 18 August, taking it out of the competition and leaving its twin without a back-up.

Meanwhile, another company, Benson Space Company in Poway, California, run by businessman Jim Benson, is looking at a hybrid approach — a rocketplane that takes off vertically but glides back horizontally. A survey by Incredible Adventures has shown that most potential passengers would prefer a horizontal landing of this sort, and Benson says that it will provide a gentler ride. Benson's rocket design is powered by a hybrid engine, which combines solid fuel with a liquid oxidizer. That gives it some of the stability of solid rockets, the motors of which can be stored fully fuelled for long periods, but with the controllability of liquid engines, which can be turned on and off during flight. The craft uses six hybrid motors, but needs only three to take off successfully. Parts of the hybrid engine technology developed by SpaceDev — another company founded by Benson — have already been used in SpaceShipOne, although just how much it contributed has been a bone of contention between Benson and Burt Rutan, the fabled aerospace engineer who created SpaceShipOne.

Under pressure

Another company building on previous success with rocket motors to plan a foothold in the spaceplane business is Xcor Aerospace of Mojave, California. Xcor already makes a small profit by selling rocket engines and low-altitude racing rocketplanes to other companies, and was included on *Inc Magazine's* list of the 500 fastest growing companies in the United States this August. It has plans for a simple spaceplane that would use a cluster of four such engines to take off from its runway and ascend to space.

Xcor has avoided finalizing the details of the design or timetable for its Xerus passenger



ROCKETPLANE KISTLER

Recycling: Rocketplane Kistler hopes that its Russian engines will be fully reusable.

rocket, but it is expected to be lightweight and small, with room for just one passenger and a pilot. "We don't want to feel rushed to market by tying ourselves to particular deadlines," says the company's president, Jeff Greason. "We never anticipated that we'd be the first to market." The company is also planning for a suborbital research market, in which their passenger spaceship could be reconfigured as a space laboratory for short-duration scientific research — the kinds of projects in astronomy, environmental monitoring, materials research and so on that are currently done on sounding rockets, which are expensive. "We've already got a customer [the Southwest Research Institute in Boulder, Colorado] waiting in the wings that wants more flights than we had thought the entire market would be," says Greason.

With these rivals at various stages of development, many are watching to see what is going on with Scaled Composites. The explosion in July was clearly a blow, although Rutan has said the company will forge ahead. Investigations into the accident are ongoing, and formal findings have yet to be released. The other major development in the company's fortunes is a buy-out by aerospace giant Northrop Grumman, which in July bought the 60% of the company that it did not already own. The price has not been disclosed, and the deal is still waiting for regulatory approval. But John Pike, the space and defence analyst who runs Globalsecurity.org, suggests that it might have had less to do with the space-tourism business than with the other projects Scaled Composites works on. The company has several patents and designs, including pilotless aircraft, that Northrop Grumman could benefit from. "It has basically bought a skunk works," says Pike, referring to the secretive, advanced-design companies pioneered by Lockheed Martin's famed 'skunk works' in California. "That kind of agility is hard to develop, so you buy it." It's even conceivable, he speculates, that Northrop could float off the personal spaceflight business as a separate company.

Growth spurt

Rick Tumlinson, cofounder of the Space Frontier Foundation in Nyack, New York, says that the acquisition could signal that traditional businesses are becoming more familiar and



Different directions: unlike SpaceShipOne's (above, with its carrier) horizontal positioning, New Shepard takes off and lands vertically.



respectful of the fledgling space-tourism industry. But he worries that a full buy-out of Scaled Composites could dampen some of its entrepreneurial spirit. "When a big aerospace company comes in and takes over one of these startups before the frontiers have been opened," he says, there's a risk that "the old-school mentality will suffocate [the new industry] even before it gets born."

Tumlinson and other advocates of a new approach to spaceflight are also concerned about the other big player entering the field: Astrium, part of the European Aeronautic Defense and Space Company. Astrium announced in June that it hopes to raise €1 billion (\$1.3 billion) to develop a rocketplane that would take off and land on a runway. It would use a single vehicle with two propulsion systems, conventional jet engines to take off and climb to the cruising altitude, and a rocket to soar into space. Charles Lurio, an aerospace engineer based in Boston, Massachusetts, and publisher of an insider newsletter on the new-space movement, says the announcement shows that "the field's become serious enough that even a company such as Astrium feels like it has to offer a competing vision."

Astrium concedes that its plans depend on raising outside funding, hopefully from private loans and regional development funding, but says that it is serious and expects to go ahead. Jeremy Close, the company spokesman, says

that its technology could also be developed into vehicles that could go beyond the space tourism market. "There's a possibility of point-to-point travel — for example, from Europe to Australia at significantly reduced travel times," he said.

The extra mile

The fact that supersonic passenger planes never got farther than the massively subsidized Concorde project of the 1960s suggests there might not be much of a civilian market for such things. But if Astrium really does manage to develop such a vehicle, it will probably catch the attention of the armed forces. The US

military, for instance, plans to use scramjet engines, which use jetplane-like intakes to extract oxygen from the air to ignite their fuel, rather than carrying a supply of oxidizer as rockets do. Codenamed Blackswift, the planes being studied would be unmanned

fighter-like craft that can reach any point on Earth within a few hours. Although the technological approach is different, Astrium's spaceplane, with its similar capabilities, might have military applications.

With so many businesses and technological approaches to achieving similar goals for private spaceflight, many analysts have a positive outlook for the industry. There's a good chance that the robust competition between these players will help to hone the systems and drive down the costs. And Logsdon adds, "I hope one or more of these folks will succeed. A little competition could be a good thing." ■

David Chandler is a freelance science writer in Massachusetts.

See Editorial, page 970.

"Most knowledgeable people think SpaceX has a reasonable chance of success."

— John Logsdon

Libya should stop denying scientific evidence on HIV

SIR — We welcome Libya's recent decision to commute to life imprisonment the death sentences of five Bulgarian nurses and a Palestinian medic. All six were imprisoned for eight years on false charges of deliberately infecting children with HIV in the hospital where they worked. We also applaud the subsequent decision by Bulgaria's president to pardon and release the six immediately upon their extradition to Sofia.

We cannot accept, however, the Libyan government's continued denial of the scientific evidence in this case. That denial constitutes a barrier to establishing normal relations with the international medical and scientific community, from which assistance is urgently needed to upgrade Libya's health-care system.

The 17 July announcement that Libya's Higher Judicial Council had commuted the death sentence to life imprisonment should have been accompanied by an explicit acknowledgement that the real cause of the outbreak was an accident stemming from insufficient infection controls and hospital safety precautions.

Indeed, the statements from Libya's prime minister and foreign minister, condemning the recent pardon, reiterating the original conspiracy charges and calling for the health workers to be re-imprisoned, shows that the final judicial and political decision-making process had little to do with the accumulated scientific evidence.

Despite pleas from more than 100 Nobel laureates (R. J. Roberts *et al. Nature* **444**, 146; 2006), the judiciary refused to allow independent scientific evidence to be heard during the trial.

In particular, the judiciary failed to take into account convincing findings that the HIV infection was present in the hospital before the arrival of the health-care workers (T. de Oliveira *et al. Nature* **444**, 836–837; 2006). Similarly ignored was the concurrent outbreak of hepatitis C among the same population of children — a strong signal that they were picking up other blood-borne infections from the hospital.

Opportunities for contamination of medical materials are frequent in many hospitals in developing countries, and there is an urgent need to redress this situation by improving health-care policies and practices. The Libyan government wants to upgrade the standards of hygiene and care in its hospitals. With the cooperation of the international scientific and medical community, it could make Libyan hospitals a model for health care in the region and in the African continent.

Removing the obstacle of Libya's intransigence on the science in this case is

essential to allow such cooperation to move forward. We call on the Libyan authorities to put this affair behind them, and to exonerate the six health-care workers.

Vittorio Colizzi*, **Tulio de Oliveira†**,
Richard J. Roberts‡

*University of Rome Tor Vergata, Via della Ricerca Scientifica, 00133 Rome, Italy

†South African National Bioinformatics Institute, University of Western Cape, Private Bag X17, Bellville 7535, South Africa

‡New England Biolabs, 240 County Road, Ipswich, Massachusetts 01938-2723, USA

This letter was also signed by:

Massimo Amicosante University of Rome Tor Vergata, Italy

Thomas Lehner Guy's Hospital, London, UK

Luc Montagnier World AIDS Foundation for Research and Prevention, Paris, France

David Pauza University of Maryland, Baltimore, USA

Luc Perrin University Hospital, Geneva, Switzerland

Giovanni Rezza Istituto Superiore di Sanità, Rome, Italy

Fires and climate linked in nineteenth century

SIR — 'Atmospheric brown clouds', resulting from the burning of fossil fuels and biomass, have recently been reported to have a large effect on climate by altering the atmosphere's absorption of solar radiation (V. Ramanathan *et al. Nature* **448**, 575–578; 2007).

Interestingly, even in the nineteenth century, some scientists held the view that tiny particles, or aerosols, produced from burning affect solar radiation, clouds and precipitation on a large scale — all factors that play into climate.

One of them, German geographer Alexander Freiherr von Danckelman, wrote an insightful but little-noticed paper on the topic (A. von Danckelman *Z. österr. Ges. Met. (Meteorol. Z.)* **19**, 301–311; 1884).

After observing huge savannah fires in Africa during the 1880s, von Danckelman reported that fires were accompanied by cumulus clouds, which subsequently spread and thinned, forming a brownish or blueish haze that persisted for days to weeks.

He argued against the view that fires were an immediate cause of rain showers, and proposed instead that they affected cloudiness and precipitation in an "indirect way". He realized that by providing cloud condensation nuclei, fires might contribute to the fog and drizzle typical of the dry season. Estimating the amount of biomass burned in Africa each year, he concluded that savannah fires must have a major influence on large-scale climate.

Von Danckelman's descriptions of haze produced from burning biomass and its effects on climate are surprisingly accurate. Although not every detail is correct, his theories anticipated many aspects of the

current discussion on biomass burning and the effects of aerosols.

Sadly his work, published in French and German, is almost forgotten today and references to his papers are absent in current studies.

Stefan Brönnimann

Institute for Atmospheric and Climate Science, ETH Zurich, Universitätsstraße 16, CH-8092 Zürich, Switzerland

Climate: Sawyer predicted rate of warming in 1972

SIR — Thirty-five years ago this week, *Nature* published a paper titled 'Man-made carbon dioxide and the "greenhouse" effect' by the eminent atmospheric scientist J. S. Sawyer (*Nature* **239**, 23–26; 1972). In four pages Sawyer summarized what was known about the role of carbon dioxide in enhancing the natural greenhouse effect, and made a remarkable prediction of the warming expected at the end of the twentieth century. He concluded that the 25% increase in atmospheric carbon dioxide predicted to occur by 2000 corresponded to an increase of 0.6 °C in world temperature.

In fact the global surface temperature rose about 0.5 °C between the early 1970s and 2000. Considering that global temperatures had, if anything, been falling in the decades leading up to the early 1970s, Sawyer's prediction of a reversal of this trend, and of the correct magnitude of the warming, is perhaps the most remarkable long-range forecast ever made.

Sawyer's review built on the work of many other scientists, including John Tyndall's in the nineteenth century (see, for example, J. Tyndall *Philos. Mag.* **22**, 169–194 and 273–285; 1861) and Guy Callender's in the mid-twentieth (for example, G. S. Callendar *Weather* **4**, 310–314; 1949). But the anniversary of his paper is a reminder that, far from being a modern preoccupation, the effects of carbon dioxide on the global climate have been recognized for many decades.

Today, improved data, models and analyses allow discussion of possible changes in numerous meteorological variables aside from those Sawyer described. Hosting such discussions, the four volumes of the Intergovernmental Panel on Climate Change 2007 assessment run to several thousand pages, with more than 400 authors and about 2,500 reviewers. Despite huge efforts, and advances in the science, the scientific consensus on the amount of global warming expected from increasing atmospheric carbon dioxide concentrations has changed little from that in Sawyer's time.

Neville Nicholls

School of Geography and Environmental Science, Monash University, Victoria 3800, Australia

COMMENTARY

Cashing in palm oil for conservation

Tropical forests in southeast Asia are under threat from oil-palm growers. This is an opportunity to combine sustainable economic growth with biodiversity conservation, argue **Lian Pin Koh** and **David S. Wilcove**.

Southeast Asia harbours 11% of the world's remaining tropical forests¹, which in recent years have been under threat from the unprecedented and explosive growth in oil palm (*Elaeis guineensis*) agriculture. The increasing global demand for products derived from palm oil, particularly in China and India, can be attributed to the oil's diverse uses. It can be used as a cooking oil, a food additive, and in cosmetics, industrial lubricants and biofuels². In some Western countries greater use of palm oil in food products is being driven by a switch away from unhealthy trans-fats. Between 1980 and 2000, the global production of palm oil increased 4.6-fold from 4.5 million to 20.9 million tonnes per year, and is projected to increase to 30.4 million tonnes per year by 2010². The rapid expansion of oil-palm crops in equatorial regions has raised concerns about its potential detrimental effects on southeast Asia's biodiversity, leading to intense media debates between environmental non-governmental organizations (NGOs) and the oil-palm industry.

NGOs contend that expansion of oil-palm agriculture in southeast Asia destroys huge tracts of tropical forests and threatens the survival of many native species, including the orang-utan (*Pongo pygmaeus*). They have launched aggressive media campaigns that lobby for the boycott of oil-palm products (see for example www.cspinet.org/palm/). In response, palm oil producers have accused Western NGOs of unfairly targeting southeast Asia's oil-palm industry while ignoring agricultural activities in other regions that also harm biodiversity, such as soya bean cultivation in South America. They also argue that oil-palm cultivation is not a threat to biodiversity because disturbed forests or existing croplands are converted to plantations with minimal disturbance to pristine habitats.

We think that this debate has been fuelled, on the one hand, by the NGOs' lack of awareness of the socioeconomic realities in countries that produce palm oil, and, on the other, by the crop growers' failure to appreciate both the threat to southeast Asia's unique biodiversity, and the conservation potential of non-pristine habitats. To break this agriculture–biodiversity deadlock, we suggest a new strategy of using revenue from oil-palm agriculture to fund the acquisition of land for the establishment



Housing benefit: many oil palm plantation companies provide housing for workers and their families.

of private nature reserves.

The oil palm is one of the most economically important crops in southeast Asia (see graphic below). Malaysia and Indonesia are the two largest producers of palm oil in the world, accounting for 80.5% of global production and 56.1% — 6.7 million hectares — of the oil palm-cultivated area worldwide³. In 2004, the export value of palm oil amounted to US\$6.3 billion for Malaysia and \$4.1 billion for Indonesia³, contributing to 5.6% and 1.7% of their gross national incomes respectively⁴. The two largest importers of palm oil in 2004 were China with 20.2 million tonnes, and India with 16.5 million tonnes³, most of which is used as cooking oil. If the current economic boom in China and India continues, it will be almost impossible to

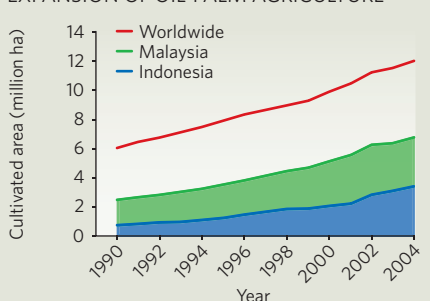
halt further expansion of the oil-palm industry in southeast Asia.

Many oil-palm plantations are effectively self-sufficient villages, providing not only employment, but also housing, basic amenities such as water and electricity, and infrastructure including roads, medical care and schools for the families of their employees⁵. As such, many communities, particularly those in rural areas, rely on oil-palm agriculture for their livelihood. According to the Malaysian Palm Oil Board in Kuala Lumpur, the country's oil-palm plantations provide direct employment for over half a million people from both the local population and neighbouring countries. Any major disruption to the palm-oil industry in southeast Asia is likely to have widespread and dire socioeconomic consequences throughout the region.

Unique biodiversity

Parts of Malaysia and Indonesia are located within two hotspots of biodiversity — Sundaland and the islands of Wallacea. These areas, which are undergoing rapid deforestation, are considered biodiversity hotspots because they contain high concentrations of species found only in these regions. For example, 89 species (44.5%) of amphibians in Malaysia and 17,500 species (59.6%) of vascular plants in Indonesia do not occur anywhere else in the world⁶.

EXPANSION OF OIL-PALM AGRICULTURE



To date, four vascular plants (*Dipterocarpus cinereus*, *Mangifera casturi*, *M. rubropetala* and *Shorea cuspidata*), one bird (*Argusianus bipunctatus*) and three mammal species (*Papagomys theodorverhoeveni*, *Paulamys naso* and *Spelaomys florensis*) in Malaysia and Indonesia have been listed as 'extinct' or 'extinct in the wild' on the World Conservation Union (IUCN) Red List⁷.

Although the number of extinct species in these regions is fairly low at this point, there are many more species facing serious threats to their survival. For example, 47 species (23.5%) of amphibians in Malaysia and 146 species (21.9%) of mammals in Indonesia are listed as threatened on the IUCN Red List⁷. Because almost all these species are unique to the region, their loss would represent global extinction.

Over the periods of 1990–2000 and 2000–2005, deforestation rates in Indonesia climbed from 2.3% to 2.7% per year for its primary (undisturbed) forests and from 1.2% to 1.3% per year for its secondary (naturally regenerated) forests⁸. The area of Malaysia's primary forests has remained unchanged at 3.8 million hectares since 1990 but the rate of loss of the country's secondary forests has increased from 0.3% per year during 1990–2000 to 0.8% per year in 2000–2005. Compared with the estimated extent of primary forests 8,000 years ago, before large-scale human disturbance, relatively little remains intact in Indonesia (25.6%) and Malaysia (11.6%)⁹.

A sustainable solution

Because Indonesia contains some three-quarters of southeast Asia's remaining primary forests, or 49 million hectares, the continuing loss of its primary forests would be disastrous for the region's biodiversity. The decline of secondary forests is also worrying. Secondary forests protected from human disturbance would eventually develop into old growth forests, which could complement the remaining primary forests in ensuring the survival of forest-dependent species. Therefore, oil-palm growers in southeast Asia need to realize that to conserve the region's unique biodiversity, it is crucial not only to protect the remaining primary forests but also to rehabilitate and preserve disturbed habitats, including logged forests and abandoned agricultural land.

Currently, less than 10% (1.5 million hectares) of the remaining tropical forests in Malaysia and just over 20% of those in Indonesia are under some form of protection by their national governments (IUCN protection categories I–VI)¹. Unfortunately, this situation is unlikely to improve much in the near future

because of the overwhelming emphasis in these countries on short-term economic growth over long-term natural resource conservation. An alternative conservation strategy adopted by some developing countries in sub-Saharan Africa and Latin America is the establishment of private nature reserves¹⁰.

Private reserves are flexible and substantial complements to traditional government-funded conservation initiatives, and have been shown



Rich pickings: crude palm oil is extracted from the yellow parts of oil palm fruit.

to be successful not only in protecting endangered species but also in providing employment for local communities¹⁰. But, owing to the scarcity of local conservation resources in developing countries, the success of this approach typically hinges on the participation of external stakeholders and donors, such as the Global Conservation Fund of Conservation International, Arlington, Virginia.

In our view, because the oil palm is such a high yielding and lucrative crop, a unique opportunity exists for NGOs to acquire relatively small tracts of existing oil-palm plantations in Malaysia and Indonesia and use the revenue generated to establish a network of privately owned nature reserves for biodiversity conservation. For example, a typical mature oil-palm plantation in Sabah, Malaysia, generates an annual net profit of roughly \$2,000 per hectare (see 'Supplementary information'). Based on the current price of \$12,500 per hectare for oil palm-cultivated land, the capital investment could be recovered in just 6 years. After this initial period, a 5,000-hectare oil palm plantation could generate annual profits amounting to some \$10 million, which could be used to acquire 1,800 hectares of forested land annually to be set aside as private nature reserves.

NGOs would need to collaborate with large conservation donor groups to fund the initial investments and with local oil-palm companies for their expertise in running the plantations. This could be a win-win partnership between

NGOs and oil-palm companies, because NGOs would be able to protect forests using the oil-palm revenue and the companies would be able to enhance their corporate image to satisfy environmentally-conscious consumers. And, because the companies know how to acquire land, and to create and manage plantations, the NGOs would not be faced with the daunting task of starting up their own oil-palm businesses. The governments of Malaysia and

Indonesia could also offer their support — for example, by leasing government-owned plantation land to these NGOs, which would facilitate the NGO's acquisition of oil-palm plantations in an increasingly competitive market.

Through such joint ventures, NGOs could both generate the funds needed for land preservation and improve the livelihoods of local communities in southeast Asia. Furthermore, because such oil-palm plantations would be motivated mainly by conservation objectives, they could provide the industry with leadership for the sustainable production of palm oil through environmentally-friendly management practices⁵. This could also drive the development of a pre-

mium market for sustainable oil-palm products and thereby generate economic incentives for more palm-oil producers to adopt sustainable practices. Through this and other novel approaches, NGOs and oil palm producers could work together to break the agriculture-biodiversity deadlock to preserve southeast Asia's last remaining tracts of tropical forests and biodiversity for future generations. ■

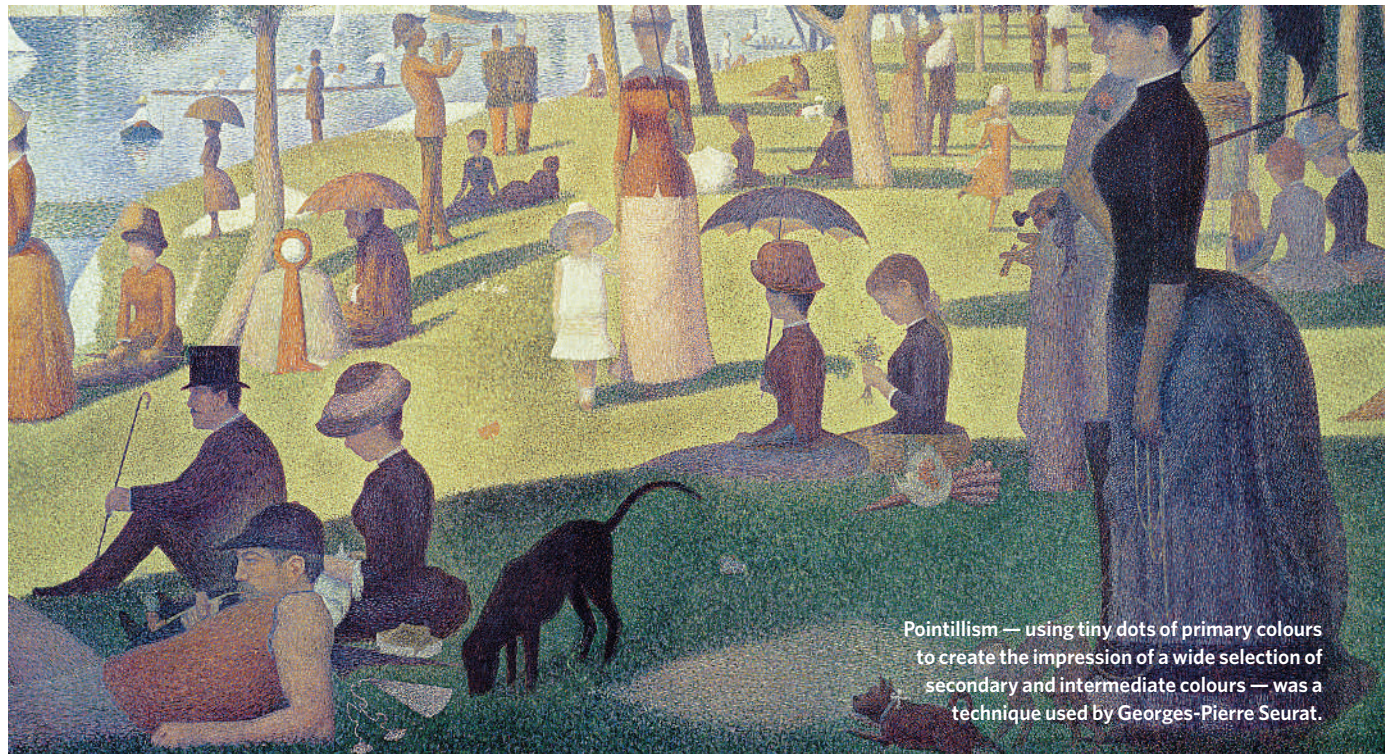
Lian Pin Koh is in the Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544, USA.

David S. Wilcove is in the Department of Ecology and Evolutionary Biology, and the Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, New Jersey 08544, USA.

1. Iremonger, S., Ravilious, C. & Quinton, T. (eds) *A Global Overview of Forest Conservation* CD-ROM, Center for International Forestry Research and World Conservation Monitoring Centre, Cambridge (1997).
2. Corley, R. H. V. & Tinker, P. B. *The Oil Palm*. Fourth Edition (Blackwell, Oxford, 2003).
3. FAOSTAT Online Statistical Service (FAO, Rome, 2006).
4. The World Bank 2006 World Development Indicators Online, www.worldbank.org/data
5. Corley, R. H. V. *Planter* **82**, 121–143 (2006).
6. UNEP-WCMC Species Data, www.unep-wcmc.org/species/data/
7. IUCN (World Conservation Union) 2006 *IUCN Red List of Threatened Species* (IUCN, Gland, Switzerland, 2006).
8. *Global Forest Resources Assessment 2005: Progress Towards Sustainable Forest Management*. FAO Forestry Paper 147 (FAO, Rome, 2005).
9. Billington, C. et al. *Estimated Original Forest Cover Map — A First Attempt* (WCMC, Cambridge, 1996).
10. Langholz, J. *Conserv. Biol.* **10**, 271 (1996).

Supplementary Information is linked to the online version of this article at www.nature.com/nature

BOOKS & ARTS



ART INST. CHICAGO/BRIDGEMAN ART LIB. NATIONALITY

An illusionary rival

Is our perception of the world a fantasy created by the brain to coincide with reality?

Making up the Mind: How the Brain Creates Our Mental World

by Chris Frith

Blackwell: 2007. 232 pp. \$24.95, £14.99

Siân Ede

In *Making up the Mind*, professor of neuropsychology Chris Frith chooses as an occasional audience a fictional professor of English whom he meets at an academic party. This “opinionated” woman “doesn’t accept that studying brain activity can tell you anything about the human mind”, he tells us. And at intervals she pops up, a sceptical adversary, embodying some generalized idea the scientist has of his literary peers, perhaps as post-modernist obscurantists or, in this case, as a vaguely unreconstructed Freudian. I am not sure how Frith (or his brain) dreamed her up, but she is the one unconvincing presence in an otherwise informative and genial book — and one longs to give her the push, or at least a plausible voice of her own.

For a start, she should already be fascinated by accounts of brain activity, possessing dog-eared copies of Oliver Sacks’ books and of Richard Gregory’s *Eye and Brain*. She would have been amused by the breezy confidence

of Steven Pinker, Susan Greenfield and V. S. Ramachandran, made politically wary by Steven Rose, and challenged by the conflicting philosophical musings of Daniel Dennett, John Searle and Galen Strawson, who regularly feature in literary journals.

António Damásio’s books would be a favoured source, and she might well be teaching her students about neuroscience as a subject for literary writing, as illustrated by Paul Broks, David Lodge and A. S. Byatt. She would have engaged with contemporary deliberations about the hard question, the illusion of self, the enigma of free will and the nature of consciousness. My guess is that she would be irritated by Chris Frith’s assumptions about her intellectual prejudices and would in fact enjoy discussing his revelations in order to apply them to her own subject.

Frith’s central thesis is that our brains create an illusion both of the world we inhabit and of our sense of personal autonomy while moving about in it — “a fantasy that coincides with reality”. Of course, the English professor would have agreed, for haven’t Kant, Wittgenstein, Merleau-Ponty and many others made similar claims, and aren’t writers and artists engaged continually with the same idea?

We use our incomplete perceptions of

the world as clues for building hypotheses about the way it feels and looks, drawing on information that is partly hard-wired, partly acquired and then recreated as plausible — or eccentric, in the case of people with brain dysfunction, or of those who are literary creations. Learning about the mechanics of the brain in receiving and remaking such impressions adds to our understanding of the creative imagination.

Frith provides plenty of striking examples. It is intriguing to learn why patients with phantom-limb syndrome or with brain damage can experience sensations and ‘visions’ that seem convincingly real to them. It is astonishing that in anyone, simply imagining an action stimulates the same brain regions that function when the action is actually undertaken.

Frith lucidly explains the mechanisms of social communication. This will strike a chord with dancers collaborating with neuroscientists on the function of mirror neurons, which are key to our ability to empathize with and mimic each other’s actions. Literary practitioners would love to understand how vivid sensations are evoked through words alone and would appreciate the stories Frith tells us about patients in the one area of scientific enquiry where a subjective account of

experience is both well documented and scientifically pertinent.

Of course, scientists need to be rigorously objective too, which is why investigations are often reduced to a narrow focus. Although our English professor might accept the elimination of contextual variables, she would be most surprised by Frith's conviction that scientific enquiry can ultimately reveal "how the world truly is" through developing increasingly plausible models of it — the subject of a section in his final chapter.

Frith draws similarities between the brain's use of partial information to devise its world view and the practice of science itself. But if this is an interpretation based on illusion, does that not suggest that truth, too, is a fabrication, and is distorted by an anthropomorphic perspective? A glimpse at the history of science shows that models of how the physical world 'works' have often been debunked. Indeed, the discipline of neuroscience itself attracts a broad range of investigators — from researchers into artificial intelligence through to quantum theorists and panpsychists — whose suppositions about what 'works', and how, may be wildly different. One doesn't need to be a cultural relativist to desire a deeper level of discussion at this point.

Frith makes an innovative, if brief, specu-

lation about the potential for a hermeneutics of neuroscience — that is, an investigation into the ways in which theories can be variously interpreted. His chapter is entitled 'How the Brain Creates Culture', but it is equally valid to ask to what extent culture creates the brain. Although some human perceptions and behaviours are undoubtedly hard-wired and autonomic, many others are acquired, not just in infancy when the brain is particularly plastic, but later in life too. The cultures we inhabit, besides the associations that we acquire as individuals, surely colour the ways in which we perceive the world — and how we form ideas about our mental processes.

Frith's approach confirms that our world picture is influenced by what Richard Gregory calls "the prevailing perceptual hypotheses", by different motivations, contexts and cognitive habits, and also, surely, by different cultural values and beliefs. This book proposes a subject that is ripe for investigation and it is one that might be productively shared by both sides, without drawing battle lines between those who are searching for one understanding of the mind and those who relish multiple explanations. ■

Siân Ede is arts director at the Gulbenkian Foundation, 98 Portland Place, London W1B 1ET. Her latest book is *Art & Science*.

personal lifestyles are concerned. Here Brown may have put himself, and his readers, into a semantic double-bind. Epidemiology is inherently preoccupied with aetiology — the investigation of the causes and origin of a disease. The fairly recent emergence of 'clinical epidemiology' does indeed concern itself with treatments and outcomes, underscoring its differentiation from the parent discipline. Moreover, for most epidemiologists the supposition that personal lifestyle risk factors take exclusionary precedence over environmental factors remains to be established.

To support his argument, Brown takes a quite narrow view of what constitutes an environmental toxic exposure, basically limiting this to either an ambient air- or waterborne anthropogenic substance. Most epidemiologists are more catholic in their approach to defining the environment. Brown, for example, dismisses indoor air exposure as a valid environmental issue, without a convincing rationale for doing so. Indeed, he is explicitly critical of epidemiologists for being overly focused on indoor air quality. But this prescriptive approach does not even take into account second-hand cigarette smoke, one example of an indoor-air environmental exposure that is highly relevant both to asthma, and, as recent data suggest, breast cancer.

Toxic Exposures does shed light on the intersection of health research, advocacy and policy-making. Participatory research features prominently in this mix, although Brown does not purport to provide methodological guidance, and it would not be reasonable to expect this from his socio-ethnographic approach to the subject. (For a practical summary of core methods, see *Social Epidemiology: Strategies for Public Health Activism* by Julie Cwikel.) Nor should Brown be criticized for a geographical focus almost exclusive to the northeastern United States, given the logistical requirements of the ethnographic methods used.

Two camps at cross-purposes

Toxic Exposures: Contested Illnesses and the Environmental Health Movement

by Phil Brown

Columbia University Press: 2007. 392 pp. \$29.50

Paul D. Blanc

Reading Phil Brown's *Toxic Exposures* caused me to revisit C. P. Snow's seminal essay, *The Two Cultures*. Snow, writing almost 50 years ago, addressed what he perceived as a dangerous gulf between the world views of scientists and non-scientists. Snow's principal concern was how educational reform might bring the two camps closer. Although Brown is narrowly focused on selected human health effects (specifically breast cancer, childhood asthma and Gulf War illnesses) in relation to environmental exposure, his book implicitly concerns the ways in which scientists and non-scientists communicate and, more to the point, how they often speak at cross-purpose.

Brown comes to his subject matter from the social sciences, relying heavily on a qualitative approach informed by ethnography. He takes theoretical guidance from systematic critiques of the health-care system, such as *Professional Dominance: The Social Structure of Medical Care* (by Eliot Freidson) and *The Social Transformation of American Medicine* (by Paul Starr). Central to the thesis of *Toxic*

Exposures is a construct that Brown labels the 'dominant epidemiological paradigm'.

Brown intends this phrase to serve as useful shorthand for a conservative biomedical mindset that should be rejected, a mindset in which disease is narrowly seen as a matter of treatment and where causation is of interest (if it is taken into account at all) only insofar as



Arsenic-contaminated water affects hundreds of thousands of people in the Indian subcontinent.

S. DAS/REUTERS

Fundamentally, however, I came away from *Toxic Exposures* confused as to its intended audience. The liberal use of didactic terms and constructs that presume fluency in the language and literature of medical sociology is likely to make the text too dense for most lay readers; there are also occasional passages that are overly doctrinaire and likely to put off even an otherwise receptive health scientist.

At one point in the text, Brown supports a seven-point manifesto of 'critical epidemiology', with the fifth point being: "Include in research reporting the explicit discussion of assumptions, values, and the social construction of scientific knowledge." This is certain to bewilder epidemiologists, most of whom see themselves as engaged scientists already bothered by a governmental policy-making

apparatus that is overly politicized in the opposite direction. Fifty years on, it seems as though the two cultures are farther apart than ever. ■ Paul D. Blanc is professor of occupational and environmental medicine at the University of California, San Francisco, California 94143-0924, USA and is the author of *How Everyday Products Make People Sick: Toxins at Home and in the Workplace*.

Celebrating four centuries of astronomy

The Telescope: Its History, Technology and Future

by Geoff Andersen

Princeton University Press: 2007. 256 pp. \$29.95

René Racine

The International Year of Astronomy in 2009 will mark 400 years of astronomical discovery, pioneered by Galileo Galilei in Padua, Italy. The anniversary will stimulate interest in the tools that made these discoveries possible, so a flush of new books on telescopes is likely to appear and sell briskly. Geoff Andersen's timely work also arrives at a point when astronomers and engineers are starting to construct a new generation of revolutionary instruments for ground-based and space-based astronomy.

Andersen's book covers the historical and technological evolution of the telescope. It is written for a broad readership and "avoids, wherever possible, the gruesome details which keep so many of us employed". It will be appreciated by laypersons and experts for its clarity and, in places, the insight it provides into the optics, engineering and, indeed, scientific purposes and limitations of telescopes.

The book is enriched by the perspective it offers on remote sensing and surveillance applications, a field that the author has contributed to as a research physicist at the US Air Force Academy. This enables him to develop some interesting speculation on the capabilities of extant 'spy telescopes'.

Historical notes are exact, lively and generally up-to-date. Witness the recent discovery of the full name of Laurent Cassegrain — seventeenth-century inventor of the Cassegrain telescope — uncovered by André Baranne and Françoise Launay in the *Journal of Optics* in 1997, and not generally known even to experts. References to current projects, such as the forthcoming Hubble Space Telescope service mission, the Thirty Meter Telescope partnership and the 100-metre-aperture OWL (for 'overwhelmingly large') telescope, are inevitably dated, given the rapid evolution of the field.

Professional astronomers and experienced amateur telescope-makers might be tempted to skip the introductory chapters on optics,



Galileo presents the most powerful telescope of its time to the Doge and Senate of Venice in 1609.

the basics of telescope design, light-analysis instruments and observing-site properties, but it would be at the cost of missing some pedagogical gems. These include the quasi-intuitive explanation of Poisson's 'bright spot', which vindicated the theory of diffraction and the wave nature of light.

Informative chapters on interferometry and advanced telescope techniques — active and adaptive optics, laser guide stars, laser communications, and so on — are seldom found in such detail in similar works, adding to the value of the book. The analogies and differences in telescopes operating in various domains, from γ -rays to radio waves and even to gravitational waves, are nicely presented and help in understanding how some more exotic instruments operate.

An author's view of a field is necessarily influenced by background and experience. Hence, Andersen's assertion that "In 1991 the US Air Force decided to declassify the existence of adaptive optics. Overnight, the field of astronomy took a giant leap forward unlike any since the development of mirrored optics" might have been couched differently by civilian French and American astronomers and by technologists who independently developed

highly productive astronomical adaptive systems in the 1980s. Likewise, some might question his estimate that "Military space technology tends to be 10–20 years ahead of that in the civilian world".

In a quite different vein, the author's opinion that "Antoine de Saint-Exupéry's accomplishments [as a surveillance pilot] are overshadowed by a children's book he wrote which is widely used to take the fun out of learning French" will sadden many who have been charmed by *Le Petit Prince* — and it is not relevant to the understanding of telescopes by laypersons.

Andersen's fine piece of work is enlightening and interesting. Experts can dispense with a few chapters that beginners will find useful, along with the appendices on mathematical notation, electromagnetic radiation and "Getting your own telescope". The text flows smoothly, enhanced by vivid discussion and pedagogical flair. And the author's take on remote sensing and surveillance confers a unique feature that will fascinate readers. ■

René Racine is professor emeritus in the Department of Physics at the Université de Montréal, PO Box 6128, Station Centre-Ville, Montréal, Québec H3C 3J7, Canada.

Linguistic gem or just another pidgin?

Talking Hands: What Sign Language Reveals about the Mind

by Margalit Fox

Simon & Schuster: 2007. 368 pp. \$27

Neil Smith

Al-Sayyid Bedouin Sign Language (ABSL) is a system of communication, created some 70 years ago by ten deaf people living in a village in Israel. In her elegantly written book, Margalit Fox claims that ABSL is “a language that is free of the influence of other languages, signed or spoken”, and uses it as the starting point for an inquiry into whether innate properties of the human mind could be reflected in the emergence of this ‘new’ language over three generations.

Fox intersperses the account of her investigation with a masterly and accessible overview of sign languages and research into them over the past half century. Although they share their linguistic properties, sign languages are independent of spoken languages. For example, American Sign Language is more closely related to French Sign Language than it is to British Sign Language, and is unrelated to spoken American English. They have phonology, expressed through the shape, location, movement and orientation of the hands, and typically exhibit a rich morphology — for example, the sign for ‘give’ can be modified to portray repeated giving, continuous giving or giving to many recipients.

Most sign languages have a complex system of verb agreement, allowing signers to keep track of who does what to whom, and to exploit word-order variation for rhetorical effect, rather than to encode grammatical relations such as subject and object. Their syntax allows the formation of questions, negations, conditionals and so on, and they have a compositional semantics. Whatever can be said in spoken language, deaf signers can convey in sign language.

The evidence for this comes from analyses of sign languages and their historical development, and from psycholinguistic and brain-imaging experiments. Parallels between the signed and the spoken in every domain are so close that it is hard to tell from a linguistic description which kind of language is being discussed. The genetically determined human faculty of language seems to be largely neutral between the two modalities.

There are differences. Signed languages are more iconic and allow for a degree of simultaneity not possible in spoken language. You can frequently tell, *post hoc*, why a sign has the shape it does, and you can sign more than one morpheme — the smallest meaningful



O. BURRIEL/SPL

Can the emergence of a new sign language tell us anything about the innate properties of the mind?

linguistic unit — at a time. Yet deviations from iconicity are frequent and signed words, like spoken words, typically consist of linear sequences of ‘locations’ and ‘movements’. The most striking difference lies in the importance of facial expression: in signed languages this can have many of the functions that grammar and intonation do in spoken languages.

Deaf people not exposed to sign languages may invent their own restricted system of communication, known as ‘home-sign’: a pidgin, which lacks many of the properties of a real language. In being passed on to succeeding generations, a pidgin may become a creole (the speaker’s or signer’s native language), and eventually turn into a full-fledged language. Pidgins have minimal morphology, creoles somewhat more, and full-fledged languages may be morphologically rich.

This progression, if it occurs without influence from other languages, is what Fox suggests can reveal innate properties of mind, and is what motivated the expedition to examine ABSL. But there are problems. Influence from other languages, both signed and spoken, has been considerable. Moreover, although pidgins are typologically identifiable, creoles are not linguistically well defined, and it is a fallacy that languages generally develop a rich morphology over time: some do (French), some do not (Chinese), and some lose much of their morphology (English).

Nevertheless, ABSL would indeed provide evidence for the operation of the human lan-

guage faculty if it really had become “without doubt a fully functioning language”. Suggestive evidence for the claim comes from the marking of grammatical relations, demonstrated syntactically by use of word order, or morphologically by use of agreement. Most sign languages rely on the latter, but ABSL chooses the former — a rigid subject–object–verb order.

Beyond this, the conclusions drawn by Fox and the team of four linguists she accompanied on the expedition are disappointing. ABSL has a large vocabulary but makes no use of spatial morphology and has no verb agreement, apparently because it is “too new”. More surprisingly, “the language seemed to lack phonology” and “holistic words serve the communicative needs of their users admirably”. So we are asked to believe that we have a ‘language’ with some (minimal) syntax but no morphology and, amazingly, no phonology.

Fox’s conclusion that “spontaneously, naturally and with no outside influence, the deaf villagers created a new human language” is overstated. ABSL seems to be not so much a language, as a pidgin that is being creolized under the pervasive influence of Israeli Sign Language.

Neil Smith is professor emeritus of linguistics at University College London, Gower Street, London WC1E 6BT, UK.

Correction

In the Book review “The art of persuasion” (*Nature* **448**, 751–752; 2007) the image erroneously portrayed Isaac Newton instead of Robert Boyle.

BEHAVIOURAL NEUROBIOLOGY

Females can also be from Mars

Nirao M. Shah and S. Marc Breedlove

Is the preference to mate as a male or a female irreversibly set during development? Apparently not: a study in mice shows that pheromone perception determines how an adult female behaves sexually.

We perceive gender as a core characteristic, generally unwavering in almost any social context. So we regard gender differences in behaviour as reflecting irrevocable, pervasive differences in the adult brain of the two sexes¹, rather than the flip of a switch between male or female behavioural repertoires. But on page 1009 of this issue, Kimchi *et al.*² suggest that, in adult female mice, two crucial components of gender — partner preference and mating behaviour — are controlled by pheromone sensing*. Startlingly, genetic or surgical disabling of pheromone perception seems to switch on full-blown male mating behaviours in females. Together with a previous study³, these experiments indicate that neural pathways responsible for male-typical sexual behaviour are present in the brains of females but lie dormant, and that it is the gender-specific processing of sensory information that determines the masculine or feminine nature of behaviour.

Pheromones are olfactory cues that aid communication of the social and reproductive status of members of a species. In vertebrates, pheromones are recognized by neurons located in two sensory tissues in the nasal cavity, the main olfactory epithelium (MOE) and the vomeronasal organ (VNO)⁴. The MOE is essential for chemoinvestigation (such as anogenital sniffing), mating and aggressive behaviour^{5,6}, whereas the VNO is required for aggressive behaviour and for identifying the sex of conspecifics^{7,8} — members of the same species.

Previous work^{7,8} had shown that deletion of the gene encoding TRPC2, a cation channel expressed only in VNO neurons, profoundly diminishes pheromone-evoked activity in these neurons. Therefore, mutant mice lacking this gene offer a highly specific means of probing the behavioural effects of diminished pheromone sensation. Male mice lacking the *Trpc2* gene do not distinguish between males and females, mating with animals of either sex^{7,8}. Moreover, in contrast to normal males, these mutant male mice do not fight with intruder males^{7,8}. Such findings had suggested

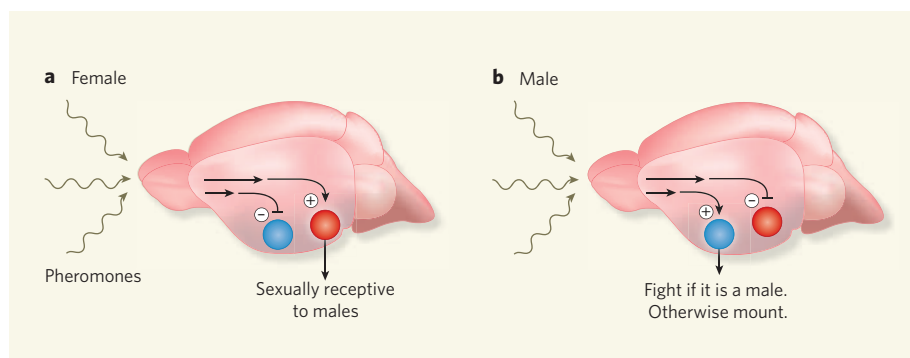


Figure 1 | A new model for the sexual differentiation of behaviour. The traditional view of sexual differentiation is that, during the development of the male brain, gonadal hormones induce the formation of neural centres controlling masculine sexual behaviours, and prevent the formation of neural centres for feminine behaviours⁹. Meanwhile, in females, feminine neural centres develop and masculine centres fail to form. Kimchi *et al.*² suggest a new model, in which neural centres controlling masculine and feminine behaviours form in both sexes during development. **a**, In adulthood, pheromonal cues in females activate the feminine sexual-behaviour centre (red) and inhibit the male centre (blue). **b**, The neural pathway for feminine sexual behaviour might also develop in male mice, but its activation is inhibited by pheromonal signals, preventing males from showing female-specific sexual behaviour. (Figure modified from Fig. 5 of ref. 2.)

that the VNO recognizes one or more male pheromones that enable gender discrimination and elicit the appropriate behavioural response. Earlier work⁷ had also shown that, unlike normal females, female mice lacking *Trpc2* do not display maternal aggression, failing to attack intruder males when nursing a litter.

Now, Kimchi *et al.*² find that *Trpc2*-deficient females also fail to distinguish between males and females among their conspecifics in terms of mating preference. Unexpectedly, however, they found that mutant females behave like *Trpc2*-deficient males, sniffing, pursuing and mounting mice of either sex. These behavioural responses do not result from a rewiring of neural circuits during development², because the authors found that normal females show similar indiscriminate, male-typical sexual behaviour when the VNO is surgically removed in adulthood. These findings suggest that the VNO detects pheromones that normally prevent female mice from displaying male-typical sexual behaviour (Fig. 1).

But *Trpc2* mutant female mice do not entirely resemble normal males in their behaviour. They also mate in a female-typical manner, bearing litters when co-housed with males.

Moreover, unlike normal males, they do not attack other males^{2,7}.

Kimchi and colleagues' findings inform our thinking about the sexual differentiation of behaviour on various fronts. Several decades of research had led to a model in which male-specific mating behaviour was thought to be hardwired into the developing brain by steroid hormones secreted from the testes⁹. The subsequent post-pubertal rise in circulating gonadal steroid hormones was thought to activate neural pathways that mediate gender-typical courtship behaviour⁹. For example, testosterone is necessary both perinatally and in adulthood to ensure male-typical sexual behaviour in rodents.

Normal adult female mice show some degree of male sexual behaviour towards other females, especially if treated with testosterone^{2,3}, when they mount nearly as often as do females lacking *Trpc2*. But Kimchi *et al.*² found that female mice lacking *Trpc2* show male-type mating patterns even in the absence of externally administered testosterone. So it seems that sexual behaviour in females is regulated both by hormonal inputs and by gender-specific neural circuits that process pheromonal cues. In

*This article and the paper concerned² were published online on 5 August 2007.

the absence of a functional VNO (as in *Trpc2*-deficient females), even low, female-typical levels of testosterone may suffice to permit high levels of male-typical behaviour.

Regardless of the mechanism involved, however, it seems that one or more male pheromones inhibit mice from mounting other male mice, and elicit aggressive behaviour in males. So when male or female mice cannot distinguish between the sexes, they initiate male-typical mating routines towards all conspecifics, presumably using pheromonal cues sensed by the MOE⁵⁻⁸ (Fig. 1).

What does the work of Kimchi *et al.*, taken together with previous findings, tell us about the functional role of structural differences in the brain? As females can display male-type mating patterns, it seems unlikely that gender-specific effectors of mating patterns lie in neural structures that differ between the sexes, such as the medial amygdala, which integrates olfactory and pheromonal information, and the medial preoptic nucleus in the hypothalamus^{1,10}. It is more likely that these sexually dimorphic regions are responsible for sex-specific changes in neural activity evoked by pheromones (Fig. 1). Such areas could also be where steroid hormones modulate the strength of male- or female-specific routines of sexual behaviour. Indeed most, if not all, of the sexually dimorphic structures in the brain express receptors for gonadal steroid hormones¹⁰. Perhaps such structures influence only those behavioural patterns, such as inter-male aggression and female sexual receptivity, that seem to be more robustly limited to one sex. Whatever the answer, the latest results suggest that sexual differentiation of neural circuits engaged by the VNO are crucial for the gender-typical display of behaviour in rodents.

Do Kimchi and colleagues' findings hold true for other organisms? Studies in the fruitfly *Drosophila* have shown that female flies genetically modified to express a male-specific messenger RNA — *fruitless^M* (*fru^M*) — engage in male-type courtship rituals with flies of either sex¹¹⁻¹³. It is not known, however, whether the *fru^M* protein confers male courtship capacity on a female when it is present only in adulthood. This protein is expressed in several sensory structures and neuronal groups in the fruitfly brain, and probably influences male courtship behaviour at many levels. Nevertheless, as Kimchi *et al.* have found in mice, sexual behaviour in fruitflies seems to be governed by simple rules, with sensory information activating either male- or female-typical mating circuits^{14,15}. Whether such simple rules dictate affairs of the human heart remains to be seen. ■

Nirao M. Shah is in the Department of Anatomy, University of California, San Francisco, 1550 4th Street, MC2722, San Francisco, California 94158, USA. S. Marc Breedlove is in the Neuroscience Program, Michigan State University, 108 Giltner Hall, East Lansing, Michigan 48824, USA.

e-mails: nms@ucsf.edu; breedsma@msu.edu

1. Morris, J. A., Jordan, C. L. & Breedlove S. M. *Nature Neurosci.* **7**, 1034–1039 (2004).
2. Kimchi, T., Xu, J. & Dulac, C. *Nature* **448**, 1009–1014 (2007).
3. Edwards, D. A. & Burge, K. G. *Horm. Behav.* **2**, 49–58 (1971).
4. Axel, R. *Sci. Am.* **273**, 154–159 (1995).
5. Mandiyan, V. S., Coats, J. K. & Shah, N. M. *Nature Neurosci.* **8**, 1660–1662 (2005).
6. Yoon, H., Enquist, L. W. & Dulac, C. *Cell* **123**, 669–682 (2005).
7. Leypold, B. G. *et al. Proc. Natl Acad. Sci. USA* **99**, 6376–6381 (2002).
8. Stowers, L., Holy, T. E., Meister, M., Dulac, C. & Koentges, G. *Science* **295**, 1493–1500 (2002).
9. Phoenix, C. H., Goy, R. W., Gerall, A. A. & Young, W. C. *Endocrinology* **65**, 369–382 (1959).
10. Shah, N. M. *et al. Neuron* **43**, 313–319 (2004).
11. Demir, E. & Dickson, B. J. *Cell* **121**, 785–794 (2005).
12. Stockinger, P. *et al. Cell* **121**, 795–807 (2005).
13. Manoli, D. S. *et al. Nature* **436**, 395–400 (2005).
14. Ejima, A. *et al. Curr. Biol.* **17**, 599–605 (2007).
15. Kurtovic, A., Widmer, A. & Dickson, B. J. *Nature* **446**, 542–546 (2007).

THEORETICAL PHYSICS

A black hole full of answers

Jan Zaanen

A facet of string theory, the currently favoured route to a 'theory of everything', might help to explain some properties of exotic matter phases — such as some peculiarities of high-temperature superconductors.

How are heat and charge transported within a high-temperature superconductor? And what happens when heavy nuclei are torn apart to make the soup of elementary particles known as a quark–gluon plasma? In a paper published on the *arXiv* preprint server, Hartnoll *et al.*¹ show convincingly that the easiest insight into the superconductor problem, just as into the quark–gluon plasma^{2,3}, is to be had by looking at a black hole. Not any old black hole, of course, but a black hole in a negatively curved space-time with an extra dimension (Fig. 1).

What might sound like a theoretical physicist's idea of a bad joke could, in fact, be history in the making. The context is a highlight of string theory known as the anti-de-Sitter space/conformal field theory correspondence⁴ — AdS/CFT for short — which demonstrates an intimate connection between Einstein's general theory of relativity and quantum physics. That it might also find use in such far-flung fields as superconductivity and the quark–gluon plasma is the stuff of physicists' dreams — the unifying power of physical laws as formulated in the language of mathematics.

Viewed as a whole, string theory amounts to a head-on attack on the incompatibility of general relativity and quantum theory, the two greatest accomplishments of twentieth-century physics. According to general relativity, space and time are dynamic entities, linked to matter and energy. By contrast, quantum physics tells us how matter and energy behave, but can only be formulated in a frozen space-time.

String theory is a collection of mathematical discoveries that might just offer a solution to this puzzle. But it has had a bad press of late. This is in part because its 40-year history is littered with claims that, if only we would stick to its true path of enlightenment, the answers to the big questions of physics would be just around the corner. Its failure to deliver on those promises and produce, so far, anything of con-

sequence to experiment has become rather an embarrassment.

The AdS/CFT correspondence is a case in point. It is a fascinating mathematical result, uncovered by the Argentinian physicist Juan Martín Maldacena in 1997, but had seemed unrelated to anything that happens in or outside the laboratory. The correspondence predicts a universe governed solely by gravity, being in this regard rather like ours, with stars, black holes and all the other familiar trappings. The difference is that it has an extra, fourth spatial dimension (plus the normal one time dimension) and a negative (anti-de-Sitter) overall curvature, so forming a universe closed in on itself.

As it turns out, this world corresponds precisely to a non-gravitating universe of just three spatial dimensions filled with something similar to the quantum fields that describe the elementary particles in the standard model of particle physics. Thus, general relativity and quantum-field theory seem to be embedded deep inside the same structure. But try as they might, string theorists have not been able to find an AdS/CFT-like theory that impinges directly on the world we live in. Until now, that is.

Hartnoll *et al.*¹ use the AdS/CFT correspondence to illuminate the real-life problem of how heat and charge currents flow in a 'quantum liquid' of electrons. These quantum liquids are found in the metallic state of copper oxide (cuprate) superconductors above the transition temperature (T_c) below which they become superconducting (which is generally around 100 kelvin). In terms of quantum mechanics, these liquids are strange beasts. Somehow, the electrons manage to organize themselves in a quantum critical state, meaning that their collective quantum physics becomes scale-invariant — it looks the same, regardless of the time- and length-scales over which one observes the system⁵. High-temperature

superconductivity below the transition temperature is commonly believed to have something to do with this enigmatic, normally conducting state, in which quantum and thermal fluctuations merge.

Although they are in all other regards vastly different, the quantum fields that pop up in the AdS/CFT correspondence share the quantum-critical (or 'conformal'; the C in CFT) property with this quantum-thermal brew. The scale invariance they share is a mighty principle; in fact, it is so powerful that the many differences between the two types of field don't matter. It was thus shown in 2001 that at energies small compared with the temperature, the transport properties of a substance containing a conformal quantum field relate mathematically through the AdS/CFT correspondence to the geometry of a black hole in anti-de Sitter space². This makes it possible to derive the equations describing transport in the quantum-thermal critical brew in a few easy lines, instead of the pages of algebra that one encounters in the direct evaluation of quantum field theory of the type that one finds, for example, in superconductivity⁶.

Hartnoll *et al.*¹ push what one might term the 'AdS-to-high- T_c correspondence' to its logical conclusion. They study its application to a particular, rather recondite transport phenomenon known as the Nernst effect — the crosswise flow of heat and charge currents in the presence of a magnetic field⁷ — in the nearly quantum-critical matter of a two-dimensional cuprate system. In a theoretical *tour de force*, they use the physics of a black hole in a three-dimensional anti-de Sitter space that carries both electrical and magnetic charge to guide them in the very complex derivation of the relevant transport equations directly from quantum field theory. They show that these theoretical results are seemingly consistent with a number of hitherto unexplained features of the Nernst effect in a high-temperature superconductor⁷.

So where does the quark–gluon plasma fit in? Here, the AdS/CFT correspondence

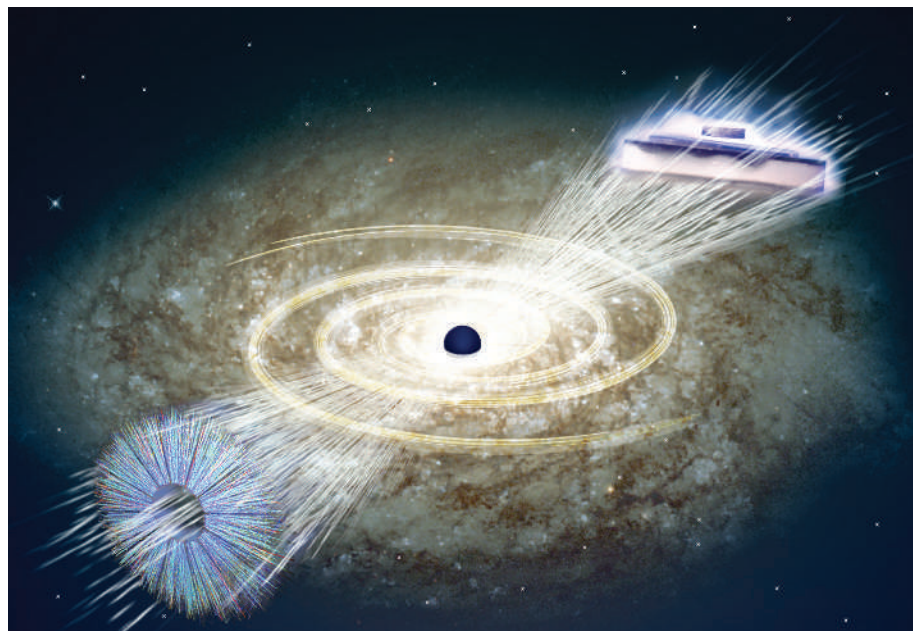


Figure 1 | The answer's out there. The easiest path to enlightenment on the mysterious phenomena of superconduction (levitating magnets; top right) and the quark–gluon plasma (bottom left; an event display from the Relativistic Heavy-Ion Collider at Brookhaven National Laboratory) leads through a black hole, say Hartnoll and colleagues¹.

comes to the aid of the experimentalists in a similar way³. The background is the observation that quark–gluon fireballs, as have been created in the Relativistic Heavy-Ion Collider at Brookhaven National Laboratory on Long Island, behave in a remarkably simple way, but one that current theories find difficult to explain — they are governed by normal hydrodynamics, but have extremely low viscosity. Quite simply, the AdS/CFT correspondence tells us that when the quantum physics is scale invariant, the viscosity of such a system can be as small as it is. This result is far from obvious given our current understanding of quantum chromodynamics, the standard-model quantum-field theory of the strong nuclear force that governs interactions in the quark–gluon plasma.

So what does this mean for the greater ambition of using string theory to unite gravity and

quantum physics? I personally take the discovery that black holes are so useful for sorting out the behaviour of real-life quantum systems as a signal that string theory might somehow be on the right track. But only further work like that of Hartnoll *et al.*¹ will confirm that hunch. ■

Jan Zaenen is at the Instituut-Lorentz, Universiteit Leiden, PO Box 9506, 2300 RA Leiden, the Netherlands.

e-mail: jan@lorentz.leidenuniv.nl

1. Hartnoll, S. A., Kovtun, P. K., Mueller, M. & Sachdev, S. Preprint at <<http://arxiv.org/abs/0706.3215>> (2007).
2. Policastro, G., Son, D. T. & Starinets, A. O. *Phys. Rev. Lett.* **87**, 081601 (2001).
3. Riordan, M. & Zajc, W. A. *Sci. Am.* **294** (5), 24–31 (2006).
4. Maldacena, J. *Sci. Am.* **293** (5), 33–39 (2005).
5. van der Marel, D. *et al. Nature* **425**, 271–274 (2003).
6. Herzog, C. P., Kovtun, P., Sachdev, S. & Son, D. T. *Phys. Rev. D* **75**, 085020 (2007).
7. Wang, Y., Li, L. & Ong, N. P. *Phys. Rev. B* **73**, 024510 (2006).

MOLECULAR BIOLOGY

Damage control

Claus M. Azzalin and Joachim Lingner

The chemical composition of normal DNA at the end of chromosomes does not differ from that of damaged and broken DNA within chromosomes. New findings hint at how the DNA-repair machinery distinguishes the two.

The maintenance of genome integrity is crucial for the survival of every organism. So even a single break along a chromosome triggers a molecular signalling cascade that leads to an appropriate DNA-damage response (DDR). This response allows recognition of the damage site and decelerates cell-cycle

progression, giving the cell a chance to repair the damage¹. Theoretically, the two free ends of each eukaryotic linear chromosome — telomeres — should evoke a similar cellular response. However, as long as they are intact, telomeres activate DDR only transiently, if at all, at defined stages of the cell cycle. In a paper

published on page 1068 of this issue, Lazzerini Denchi and de Lange² provide clues on how this is achieved at a molecular level.

Telomeres consist of serial repeats of nucleotides terminating in a 3' protruding, single-stranded sequence. Telomeric DNA associates with a six-protein complex known as shelterin³, which shelters the DNA from recognition by the DDR pathways as sites of damage. Lazzerini Denchi and de Lange show² that two of the shelterin proteins, TRF2 and POT1, independently repress the two main DDR pathways, which are normally induced by damage to DNA sequences within chromosomes.

In most cells, telomeres progressively erode as cells go through successive cycles of division; this is because of incomplete replication

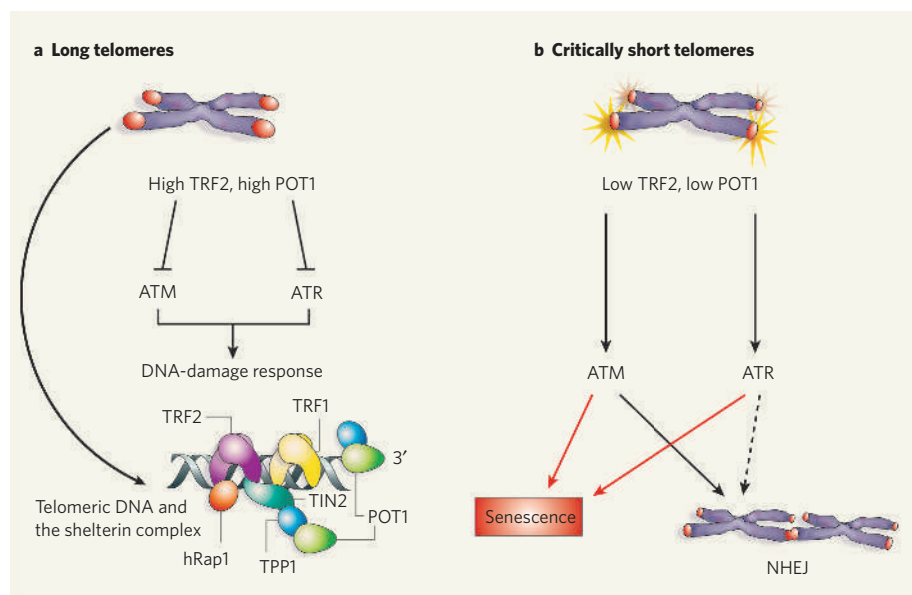


Figure 1 | Protection of telomeres from the DNA-damage response. In cells that do not express telomerase, telomeres become shorter with each cycle of cell division. **a**, Long telomeres bind to several units of the shelterin complex, which consists of six proteins: TRF1, TRF2, hRap1, TPP1, TIN2 and POT1. This complex ensures that telomere ends, which are similar in chemical composition to broken DNA sequences within chromosomes, are not mistaken for sites of DNA damage by the ATM- and ATR-mediated DNA-repair machinery. Of the shelterin proteins, TRF2 and TRF1 bind to the double-stranded telomeric DNA, whereas POT1 binds to the single-stranded telomeric 3' overhang sequence, as well as to TPP1. **b**, When telomeres become critically short, they induce a permanent arrest in the cell cycle called cellular senescence. Such short telomeres were known⁵ to activate ATM and ATR kinases, which mediate the DNA-damage response. Lazzerini Denchi and de Lange² now identify structural changes that lead to the activation of ATM and ATR at telomeres. They find that the absence of TRF2 activates ATM and the loss of POT1 triggers ATR. Loss of TRF2 also induces chromosome end-to-end fusion by the process of non-homologous end-joining (NHEJ).

of DNA ends by classical DNA polymerase enzymes, the trimming of telomere ends by nucleases, and the absence of the telomere-lengthening enzyme, telomerase⁴. On reaching a critically short length, telomeres induce a permanent arrest in the cell cycle through a process called cellular senescence, which is thought to be a powerful tumour-suppressive mechanism. Cellular senescence is triggered by the same DDR pathways that function during genuine damage⁵.

Activation of DDR relies on the functioning of one of two protein kinases, ATM and ATR, which regulate the activity of downstream DDR factors by adding phosphate groups to specific amino-acid residues¹. To investigate which of these DDR pathways TRF2 provides protection from, Lazzerini Denchi and de Lange deleted the gene encoding this protein in either ATM-deficient or ATR-depleted cells. They found that, in the absence of TRF2, DDR is activated only in cells that have normal levels of ATM, indicating that TRF2 protects telomeres from ATM-mediated DDR.

In senescent cells, DDR is activated not only through ATM-mediated pathways but also through ATR-mediated pathways⁵. So the authors asked which shelterin protein suppresses ATR activity at telomeres in these cells. They tested the role of POT1, which accumulates at telomeres both directly by binding to the single-stranded, 3' overhang DNA sequences and indirectly by interacting

with the double-stranded sequences through protein-protein interactions with other shelterin components, namely TPP1, TIN2 and TRF1 (Fig. 1a).

The authors found that, in POT1-deficient mouse cells, reducing the levels of ATR, but not ATM, leads to a decrease in accumulation of DDR factors at telomeres; they also confirmed this observation in human cells. Thus, POT1 seems to have an essential function in protecting telomeres from ATR-mediated DDR. As ATR mainly signals DNA damage caused by aberrant DNA replication¹, it will be interesting to test whether POT1 functions to ensure complete replication of telomeres.

In addition to DDR, unprotected telomeres can also trigger unwanted ligation of different, unprotected telomeres by the non-homologous end-joining repair machinery⁶. In the canonical DDR, induced by DNA double-strand breaks at intrachromosomal regions, the two free ends of each broken sequence are resealed together. Although intact telomeres are protected from this repair pathway, TRF2-depleted telomeres are fused to one another by inappropriate non-homologous end-joining, leading to circular chromosomes or long trains of ligated chromosomes⁶.

Lazzerini Denchi and de Lange found that in the absence of ATM, telomeres lacking TRF2 are not efficiently ligated to each other anymore. Moreover, when POT1 levels were reduced in these TRF2-depleted, ATM-deficient cells —

that is, when ATR has had access to the telomeres — non-homologous end-joining was triggered in a way that was largely dependent on ATR. These findings reveal unanticipated and crucial functions of ATM and ATR in the non-homologous end-joining of dysfunctional telomeres.

The experiments of Lazzerini Denchi and de Lange clearly attribute fundamental functions to mammalian TRF2 and POT1 in, respectively, repressing ATM and ATR at telomeres. The authors suggest a model in which diminished loading of TRF2 and POT1 at short and old telomeres leads to de-repression of the ATM and ATR kinases to induce cell-cycle arrest and cellular senescence (Fig. 1b).

It will be a challenge to test this model in senescent cells, because even within the same cell, telomeres vary in size, and a single, short chromosome end may be sufficient to elicit DDR. Moreover, transient ATM and ATR signalling seems also to occur at intact telomeres during or after DNA replication⁷. This does not induce senescence and is thought to be crucial for both mediating structural changes to telomeres and activating telomerase in cells that express it. Crucial functions of ATM and ATR in telomerase activation are well-established in yeast, in which ATM helps recognition of short telomeres by telomerase, promoting their preferential elongation^{8–11}, and simultaneous deletion of both ATM and ATR leads to a complete inactivation of telomerase at chromosome ends^{12,13}.

Thus, telomeres seem to adopt a variety of structures to either prevent or activate ATM- and ATR-mediated signalling, which, in turn, may either inhibit or activate cellular senescence, non-homologous end-joining or telomerase. In this structural puzzle, POT1 and TRF2 are crucial building-blocks. Elucidation of both the exact telomere composition and the structural states associated with its different possible fates remains a challenge for the future.

Claus M. Azzalin and Joachim Lingner are at the Institut Suisse de Recherche Expérimentale sur le Cancer (ISREC), 'Frontiers in Genetics' National Center for Competence in Research, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1066, Epalinges s/Lausanne, Switzerland. e-mail: joachim.lingner@isrec.ch

1. Sancar, A., Lindsey-Boltz, L. A., Unsal-Kacmaz, K. & Linn, S. *Annu. Rev. Biochem.* **73**, 39–85 (2004).
2. Lazzerini Denchi, E. & de Lange, T. *Nature* **448**, 1068–1071 (2007).
3. de Lange, T. *Genes Dev.* **19**, 2100–2110 (2005).
4. Bodnar, A. G. et al. *Science* **279**, 349–352 (1998).
5. d'Adda di Fagagna, F. et al. *Nature* **426**, 194–198 (2003).
6. van Steensel, B., Smogorzewska, A. & de Lange, T. *Cell* **92**, 401–413 (1998).
7. Verdun, R. E. & Karlseder, J. *Cell* **127**, 709–720 (2006).
8. Sabourin, M., Tuzon, C. T. & Zakian, V. A. *Mol. Cell* **27**, 550–561 (2007).
9. Bianchi, A. & Shore, D. *Genes Dev.* **21**, 1726–1730 (2007).
10. Chang, M., Arneric, M. & Lingner, J. *Genes Dev.* (in the press).
11. Arneric, M. & Lingner, J. *EMBO Rep.* (in the press).
12. Naito, T., Matsuura, A. & Ishikawa, F. *Nature Genet.* **20**, 203–206 (1998).
13. Ritchie, K. B., Mallory, J. C. & Petes, T. D. *Mol. Cell. Biol.* **19**, 6065–6075 (1999).

PLANETS

The first movement

Jeff Cuzzi

How do large objects form from the dusty gas surrounding a young star? A simulation suggests that several familiar processes, among them gas turbulence and self-gravitation, might work together to get the job done.

Making planets is tricky, and probably takes several stages. First, tiny interstellar grains must accrete into mountain-sized objects massive enough to decouple from their cocoon of nebula gas. These objects probably then combine in collisions, growing ever larger, past asteroid-sized planetesimals and lunar-sized embryos, to full-blown planets. How the first stage of this process, primary accretion, works is a fundamental unsolved problem of planetary science. On page 1022 of this issue, Johansen *et al.*¹ show how a combination of previously studied processes, acting together, might be the answer.

Our understanding of how protoplanetary nebulae evolve is generally based on observations of regions where stars are forming today. But the domain near the midplane of a nebula, where large objects grow, is shrouded from observations at visual and infrared wavelengths by opaque dust at higher altitudes in the nebula. And for longer-wavelength studies, insufficient spatial resolution is a problem.

Fortunately, our Solar System provides us with actual samples of primary planetesimals, in the form of primitive meteorites from asteroids and, recently, a milligram of cometary material returned by NASA's Stardust mission². These planetesimals consist mainly of millimetre-sized particles — silicate 'chondrules' and higher-temperature oxides — often individually melted by intense thermal pulses in the nebula³. The ages of these sand-sized grains, assessed from a growing body of radioisotope data, indicate that primary accretion was an inefficient process that took between 1 million and 3 million years⁴.

Over such a long period, according to models, the density, temperature and composition of the nebula would have changed profoundly⁵. Centimetre-to-metre-sized particles would also have migrated long distances, redistributing the nebula's solid component relative to its gas⁶. The mineral composition of the particles changed with their environment, and the result was the pot-pourri of meteorite classes with differing ages, structures, chemistry and isotopic content that we see today. Working backwards from today's evidence to infer the environment and physics of the primary accretion process is a fascinating challenge.

Take turbulence, for instance. Tiny dust grains routinely seen floating far above the midplane of million-year-old protoplanetary disks beyond our Solar System, and crystalline silicate grains seen in abundance in cometary nuclei², can be explained if nebula turbulence transports them around. But just what process can provide

the energy needed to maintain turbulent gas motions, which would be quickly damped by the viscosity of the gas, remains controversial.

One popular mechanism is magnetorotational instability, in which low-density, ionized nebula gas couples to ambient magnetic fields strongly enough that tiny velocity fluctuations are amplified⁷. It is unclear whether this type of turbulence can exist in the dense, neutral gas found in the inner regions of protoplanetary nebulae, but other types of instability might occur even if the nebula gas is not ionized (ref. 8 and references therein). Once a turbulent regime forms, its details are only weakly dependent on the driving process, much as was suggested by the Soviet mathematician Andrei Kolmogorov more than 60 years ago⁹.

Johansen *et al.*¹ construct a computational model of a nebula in which magnetorotational instability (for example) drives realistic three-dimensional turbulence. They then follow the evolution of metre-sized boulders in the model. The turbulence induces relative velocities between particles, and leads to local fluctuations in gas pressure, often on fairly large scales. These pressure fluctuations affect the local gas velocity, and determine whether particles experience a head wind (and drift inwards towards the central star) or a tail wind (and drift outwards). Particles thus accumulate in radial pressure maxima, which come and go with the evolution of large, turbulent eddies. Metre-sized particles tend to accumulate the most, as their drift rate under pressure fluctuations is the fastest.

Some of these zones live long enough that the rapidly drifting, metre-sized particles can become dense enough to exceed the gas density. In this case, the particles accelerate the surrounding gas to their preferred (keplerian) orbital velocity. These dense clumps sweep up other particles, and so become ever more massive. Ultimately, some clumps become bound together by their own gravity and, presumably, remain so as the components of the clump migrate slowly towards their mutual centre of gravity and become dense rubble piles. In only a few orbit periods, the model sees large bound clumps emerging, some as massive as the largest body in the asteroid belt, the 900-kilometre-diameter Ceres.

Johansen *et al.*¹ are appropriately cautious about some remaining uncertainties, particularly several poorly understood 'geological' issues related to the mechanics of solids. Physical sticking is needed to create metre-sized particles initially, and inelastic collisions (in which kinetic

energy is converted into other forms of energy) are needed to damp their random velocities.

Sticking seems likely for small particles because they collide at very low velocities¹⁰. But collisional velocities in turbulence may reach tens of metres per second for metre-sized particles^{11,12}, a value more likely to destroy the particles than to bind them together. Frequent disruption at this crucial stage would weaken the drift-concentration effect that provides the catalyst for growth in this model. Because of limited numerical resolution and uncertainties in the physical properties of realistic nebula particles, the results of this paper are preliminary in this regard, as the authors describe in their extensive online supplementary information. Both improved numerical models of the coupled dynamics of gas and particles and a better understanding of the mechanical properties of aggregates¹³ are needed.

The processes described by Johansen *et al.* emphasize concentration of metre-sized boulders, which themselves grew merely by the sticking of smaller grains¹⁰. Thus, one of the longest-known aspects of primitive meteorites, their dominance by millimetre-scale particles that are highly sorted according to size³, is left unexplained. A different kind of concentration process has been proposed to explain how millimetre-sized particles can be concentrated into dense zones in the nebula¹², but so far only qualitative ideas of how planetesimals might grow out of these zones have been advanced.

The answer could be that some combination of processes, each selecting a different particle size, acts simultaneously or sequentially, possibly in turbulent conditions. (Of course, the mechanism by which turbulence is maintained remains uncertain.) Whatever the final answer turns out to be, the results of Johansen and colleagues¹ indicate that future efforts devoted to developing more complex models of the interactions between particles and gas in the protoplanetary nebula will be a good investment. ■

Jeff Cuzzi is at the NASA Ames Research Center, Moffett Field, California 94035, USA.
e-mail: jcuuzzi@arc.nasa.gov

1. Johansen, A. *et al.* *Nature* **448**, 1022–1025 (2007).
2. Brownlee, D. P. *et al.* *Science* **314**, 1711–1716 (2006).
3. Scott, E. R. D. & Krot, A. N. *Astron. Soc. Pac. Conf. Series* **341**, 15–54 (2005).
4. Kita, N. *et al.* *Astron. Soc. Pac. Conf. Series* **341**, 558–587 (2005).
5. D'Alessio, P., Calvet, N. & Woolum, D. *Astron. Soc. Pac. Conf. Series* **341**, 353–372 (2005).
6. Ciesla, F. & Cuzzi, J. N. *Icarus* **181**, 178–204 (2006).
7. Stone, J. M., Gammie, C. F., Balbus, S. A. & Hawley, J. F. in *Protostars and Planets IV* (eds Mannings, V., Boss, A. P. & Russell, S. S.) 589–611 (Univ. Arizona Press, Tucson, 2000).
8. Mukhopadhyay, B. *Astrophys. J.* **653**, 503–512 (2006).
9. Kolmogorov, A. N. *Dokl. Akad. Nauk. SSR* **30**, 301–305 (1941).
10. Dominik, C., Blum, J., Cuzzi, J. & Wurm, G. in *Protostars and Planets V* (eds Reipurth, B., Jewitt, J. & Keil, K.) 783–800 (Univ. Arizona Press, Tucson, 2000).
11. Weidenschilling, S. J. in *Meteorites and the Early Solar System* (eds Kerridge, J. F. & Matthews, M. S.) 348–371 (Univ. Arizona Press, Tucson, 1988).
12. Cuzzi, J. N. & Weidenschilling, S. J. in *Meteorites and the Early Solar System II* (eds Lauretta, D. & McSween, H.) 353–382 (Univ. Arizona Press, Tucson, 2006).
13. Sirono, S. *Icarus* **167**, 431–452 (2004).

TRANSLATION

Duality in the genetic code

John F. Atkins and Pavel V. Baranov

The encoding of two non-universal amino acids involves dynamic redefinition of 'stop' signals in the genetic code. Bacteria with multiple proteins containing these amino acids add to our appreciation of coding versatility.

One of the greatest achievements of twentieth-century biology was the deciphering of the genetic code in the mid-1960s. This feat centred on the discovery that the nucleotides of messenger RNA containing the uracil, adenine, guanine and cytosine bases (abbreviated to U, A, G and C, respectively), are decoded in triplets, or codons, to signify individual amino acids and the 'stop' codons that terminate the synthesis of a protein.

Since then, the common perception has been that each triplet in the genetic code has an unambiguous meaning, coding for one of the 20 common amino acids or for stop. However, two additional amino acids — selenocysteine and pyrrolysine — are now known to be directly encoded in a tiny number of genes. They are specified by what are stop codons in essentially all other genes, and interpreting these codons involves a dynamic competition with the termination function. Writing in *Nucleic Acids Research*, Zhang and Gladyshev¹ report the discovery of a bacterium that has an unusually high number of proteins containing these two amino acids. Their paper highlights the increasing awareness of the flexibility of the genetic code^{2–5}.

A changed meaning of a triplet codon does not necessarily involve additional amino acids or dual meaning. The new meaning can be independent of any particular mRNA or location within a coding sequence. This context-independent, total reassignment of particular codons (Fig. 1a) occurs mostly in specialized niches, such as mitochondria.

Alternatively, the redefined meaning can be context-dependent — that is, apply only in a designated mRNA, or even just at a site within an mRNA. Whereas the latter pertains when the stop codons UGA and UAG are redefined to specify certain common amino acids, the encoding of selenocysteine by the stop codon UGA exemplifies both attributes⁶. In bacteria there is a site-specific recoding signal embedded in mRNA that flags a nearby UGA codon as encoding selenocysteine rather than stop (Fig. 1b). In other organisms, ranging from worms to mammals, a recoding signal located outside the coding part of the mRNA (Fig. 1c) signifies that all UGA codons in its mRNA specify selenocysteine (one essential human mRNA has ten UGAs).

Pyrrolysine, which is encoded by the stop codon UAG, presents a different case. In

pyrrolysine-containing organisms other than bacteria, a strategy intermediate between total reassignment and context-dependent redefinition is used. As far as is known, an embedded signal is not always present in these organisms (a subset of archaea). But where it does occur, it enhances the efficiency of a nearby UAG in encoding pyrrolysine — in other words, it aids the process, but is not required⁷. Because UAG, with at least some efficiency, specifies pyrrolysine without any mRNA signal, it might be expected that UAG would not be the terminator at the ends of coding sequences in these organisms. Indeed, in pyrrolysine-encoding archaea, UAG occurs only very rarely at such a location, and when it does, a non-UAG stop codon is nearby.

Until now, only one bacterial species was known to have the machinery to encode pyrrolysine. Although this bacterium apparently has only three UAGs at internal positions encoding pyrrolysine, it has UAG at the ends of thousands of coding sequences, where it probably has to function as an effective stop signal. In this case, one might expect something to flag the three UAGs in the middle of the relevant coding sequences as being special; or conceivably there could be an element discriminating against pyrrolysine specification at other UAGs. But the situation remains mysterious.

Zhang and Gladyshev's finding¹ may help address this mystery. They have looked at a 'metagenomic' data set representing bacteria that live symbiotically with an annelid worm. These bacteria live just below the outer skin of the worm, which lacks a mouth, a gut and nephridia⁸, and inhabits sediments off the coast of Italy. The symbionts collectively provide nutrients for their host and take care of its nitrogenous waste.

The authors' analysis reveals that one of the bacteria — delta-proteobacterium $\delta 1$ — has eight genes that have internal UAG codons and encode proteins expected to contain pyrrolysine. This significant expansion of the known repertoire of pyrrolysine-coding proteins will permit tests for mRNA contextual signals. Such tests might include progressive deletions of the gene containing the pyrrolysine-specifying codon in a genetically tractable organism to which the pyrrolysine-insertion machinery has already been transferred⁹.

The $\delta 1$ symbiont also has the largest known number of proteins containing selenocysteine — 57, more than twice as many as in humans. Thirteen of these are new to science¹, and their study will extend our knowledge of how the unique properties of selenocysteine can be exploited.

Research on expanding the genetic code artificially may be guided by investigation of how the $\delta 1$ symbiont translational machinery solves the problem of potentially ambiguous genetic readout when decoding codons with dual meaning. There is considerable interest in developing technologies for incorporating novel amino acids with beneficial properties

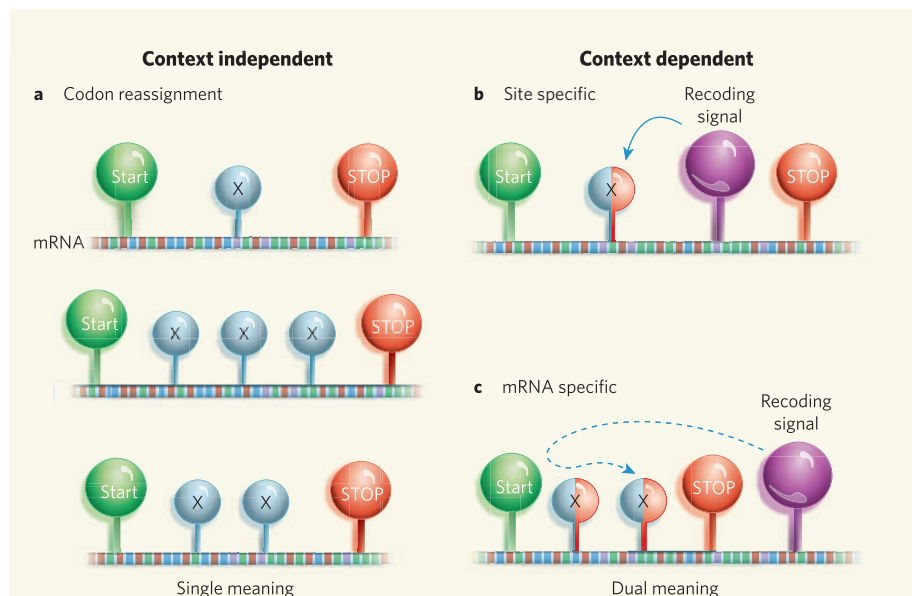


Figure 1 | Changing the meaning of codon translation. **a**, Stop codons (in reassigned form denoted as X) can be reassigned independently of the context wherever they occur in any messenger RNA in a cell. Three example mRNA sequences are shown. This mode of changing codon meaning occurs in certain specialized genomes, such as those of many mitochondria. **b** and **c**, In context-dependent reassignment, either **(b)** site-specific codon redefinition can occur in response to a co-localized recoding signal (shown as a loop representing an RNA structure), or **(c)** mRNA-dependent redefinition of codons might happen in response to a recoding signal embedded outside the coding region.

in cellular proteins. Parallels are emerging between known aspects of the natural mechanisms and developing synthetic approaches. Rather than redesigning small, whole genomes to make certain codons available for specifying novel amino acids, one approach has been to use stop codons and redefine their meaning — which is conceptually similar to both strategies used naturally in selenocysteine encoding (Figs 1b, c). In such work¹⁰, a core component of the decoding apparatus, the ribosome, was modified to tip the balance against termination. The components of the ribosomes responsible for recognition of translation start sites were also modified in parallel with the start sites themselves. This ensures that the doubly altered ribosomes initiate translation only on designated mRNAs containing the codon specifying the novel amino acid. ■

John F. Atkins is at the BioSciences Institute, University College Cork, Cork, Ireland, and the

Department of Human Genetics, University of Utah, Utah 84112-5330, USA. Pavel V. Baranov is at the Department of Biochemistry, University College Cork, Cork, Ireland.
e-mails: john.atkins@genetics.utah.edu;
p.baranov@ucc.ie

1. Zhang, Y. & Gladyshev, V. N. *Nucleic Acids Res.* doi:10.1093/nar/gkm514 (2007).
2. Yarus, M., Caporaso, J. G. & Knight, R. *Annu. Rev. Biochem.* **74**, 179–198 (2005).
3. Vetsigian, K., Woese, C. & Goldenfeld, N. *Proc. Natl Acad. Sci. USA* **103**, 10696–10701 (2006).
4. Ambrogelly, A. et al. *Nature Chem. Biol.* **3**, 29–35 (2007).
5. Namy, O., Rousset, J. P., Naphthine, S. & Brierley, I. *Mol. Cell* **13**, 157–168 (2004).
6. Hatfield, D. L., Berry, M. J. & Gladyshev, V. N. (Eds) *Selenium: Its Molecular Biology and Role in Human Health* (Springer, New York, 2006).
7. Longstaff, D. G. et al. *Mol. Microbiol.* **63**, 229–241 (2007).
8. Woyke, T. et al. *Nature* **443**, 950–955 (2006).
9. Longstaff, D. G. et al. *Proc. Natl Acad. Sci. USA* **104**, 1021–1026 (2007).
10. Wang, K., Neumann, H., Peak-Chew, S. Y. & Chin, J. W. *Nature Biotechnol.* **25**, 770–777 (2007).

PALAEOCLIMATE

Oxygen's rise reduced

Timothy W. Lyons

Why did oxygen not appear in Earth's atmosphere until hundreds of millions of years after photosynthesizing organisms first produced it? Perhaps because reducing gases from undersea volcanoes claimed it first.



The emergence of volcanoes from the sea might have been a crucial step in establishing an oxygen-rich atmosphere on Earth.

The oxygen in Earth's atmosphere is almost exclusively a product of photosynthesis. The transition from an early, virtually oxygen-free world to an irreversibly oxygenated one is linked to the first appearance and proliferation of photosynthesizing cyanobacteria. But whereas the first notable trace of persistent atmospheric oxygen has been dated to around 2.4 billion years ago^{1,2}, the fingerprints of cyanobacteria seem to stretch back as much as 2.7 billion years³, if not further.

To explain this time lag, Kump and Barley⁴ (page 1033) argue that reactions between oxygen and reduced volcanic gases initially kept oxygen levels low. This mechanism was later compromised when continents formed and stabilized: as volcanoes moved from the sea

floor to the land, the gases they produced would have been less reducing, and so would have lost much of their capacity to consume oxygen.

The timing of oxygen's rise is pinpointed most convincingly by sulphur-isotope ratios preserved in sedimentary minerals such as pyrites (formed from iron sulphide, FeS₂)^{1,2}. There are four non-radioactive sulphur isotopes, and these participate to differing extents in most biological, physical and chemical processes in a predictable way that is related to isotope mass. Unusually, however, their degree of participation in atmospheric reactions catalysed by ultraviolet light is independent of their differing masses.

Molecular oxygen (O₂) in the atmosphere erases the unique signatures of these reactions,

and also forms ozone that stifles the penetration of ultraviolet radiation into the atmosphere. It is thus the enemy of both the production and the preservation of mass-independent sulphur fractionations. The last appearance of such fractionations in pyrites can be dated to about 2.4 billion years ago, and so this point marks a clear shift in atmospheric chemistry to a state more abundant in oxygen⁵.

But what caused this shift? Most models of the early atmosphere predict appreciable, relatively constant oxygen production during the roughly 300 million years before atmospheric oxygenation. They therefore rely on vast and efficient sinks that consumed oxygen as fast as it was produced. A shift to a more oxidizing atmosphere would require a loss in the biosphere's ability to prevent the accumulation of oxygen.

One possible mechanism is a spurt of biomass burial at sea. During the short-term cycling of carbon in the biosphere, the decay of organic remains consumes much of the oxygen released during photosynthesis. Long-term burial of this organic matter as it settles to the sea floor steals food from carbon-respiring microbes, and so shifts the balance away from their oxygen-consuming metabolisms. Overall oxygen levels increase — a process that would be expressed in changes in the isotopic composition of carbon dissolved in the ocean and preserved in limestones precipitated in equilibrium with seawater. The carbon-isotope record points to abundant organic burial in the early ocean. But many researchers, among them Kump and Barley⁴, argue that the lack of any significant shift in the isotope data discounts a sudden upsurge in the amount buried as the cause for the change in oxygen levels 2.4 billion years ago⁶.

An alternative explanation is the loss of a 'buffering' capacity that held oxygen and prevented its release into the atmosphere. A range of possible buffers existed, including the reduced form of iron, Fe²⁺, which was abundant in the early ocean largely because of the input of high-temperature fluids from vents on the sea floor. Great volumes of Fe³⁺-rich banded iron formations (BIFs) peaking at 2.5 billion years ago testify to the oxidation of Fe²⁺. Yet despite their abundance (which makes BIFs of vast economic importance as sources of iron), the iron sink for oxygen is not enough to explain atmospheric trends.

Building on past work^{6–8}, Kump and Barley⁴ explore the importance for buffering via reduced gases such as hydrogen (H₂) and hydrogen sulphide (H₂S). These are most commonly released from volcanoes on the sea floor, and consume oxygen as they react to form water and sulphate (SO₄^{2–}), respectively. The amount of oxygen thus consumed is strongly dependent on the temperature of the volcanic eruptions, with the lower temperatures of the sea floor resulting in greater release of the reduced gases.

The authors compiled a large data set on

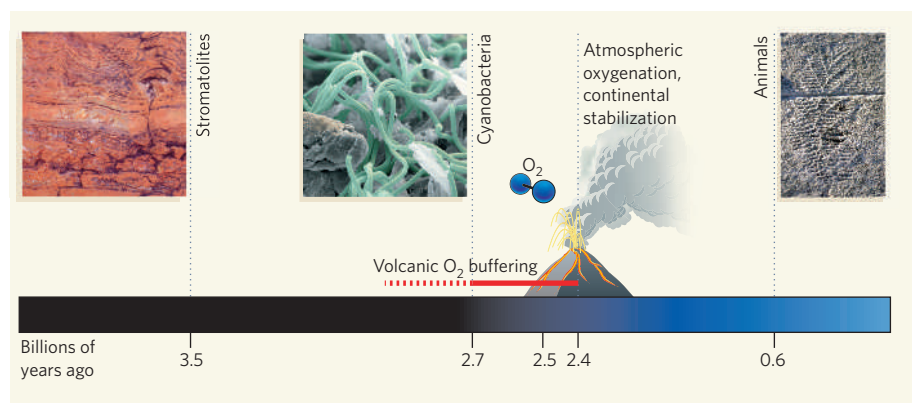


Figure 1 | Oxygen on the up. A gap of around 300 million years separates our best estimate for the first significant production of oxygen by cyanobacteria 2.7 billion years ago (Gya) and irrefutable evidence for the rise in atmospheric oxygen around 2.4 Gya. Some estimates put the first appearance of cyanobacteria back to 3.5 Gya or more, bolstered by the appearance of stromatolites — fossilized microbial mats — in the geological record, but a cyanobacterial origin for these earliest mats has not been demonstrated convincingly. Kump and Barley⁴ explain the time lag by invoking reduced gases released from abundant submarine volcanism up until 2.5 Gya, which consumed oxygen produced photosynthetically by cyanobacteria. As continents stabilized, the locus of volcanism shifted towards the land, gas emanations became more oxidized, and the buffer that inhibited oxygen build-up waned. After around 2.4 Gya, atmospheric oxygen began its climb towards levels that, almost 2 billion years later, spawned the rise of animals.

volcanic activity through time and show that the abundance of submarine volcanoes “abruptly and permanently diminished” at the boundary between the Archaean and Proterozoic eras — defined at 2.5 billion years ago. At the same time, eruptions on land became more common as continents became larger and more stable through the amalgamation of earlier continental fragments. The oxidized gases such as carbon dioxide that dominate these eruptions at higher temperatures are much less able to consume atmospheric oxygen. The timing of these events is more than coincidence, suggest Kump and Barley⁴.

But the strange case of the delayed rise of atmospheric oxygen isn't closed. The 2.7-billion-year-old biomarker record, although widely favoured, is not beyond challenge, with some researchers placing the onset of cyanobacterial production later, closer in time to the initial atmospheric oxygenation⁹, and thus eliminating the problem highlighted by Kump and Barley. Others dispute the assumptions of high and consistent carbon burial during the Archaean¹⁰. Still others¹¹ posit that oxygenic photosynthesis began more than 3.7 billion years ago, basing this on carbon-isotope data that could point to oxygen production and inferences about early cycling of uranium in the ocean.

If these wrinkles in the story were not enough, the sulphur isotopes tell us only that the ‘great oxidation event’⁶ 2.4 billion years ago demanded an increase in atmospheric oxygen content from less than 0.001% to slightly more than 0.001% of the present level⁵, although greater change is possible. Unsurprisingly, given the sensitive balance between sources and sinks, oxygen content might have vacillated before its irreversible rise¹², yet recent claims of temporary increases in atmospheric oxygen before the biomarker record for

cyanobacteria^{4,13} at 2.7 billion years ago are not persuasive¹⁴. Oxygen remained at only a small percentage of today's atmospheric and oceanic levels¹⁵ until the late Proterozoic era almost 600 million years ago¹⁶ — the time that saw the rise of the earliest animals (Fig. 1).

Despite the long march towards an oxygen-rich world, it doesn't pay to be too unimpressed by such small and perhaps geologically rapid atmospheric changes. Crossing the threshold 2.4 billion years ago triggered oxidative

continental weathering, Earth's first widespread glaciation in the wake of increased methane oxidation, and compositional changes in the ocean that redefined the ecosystems of prokaryotes (those organisms, the bacteria and archaea, that lack a discrete membrane-bounded cell nucleus), if not of their further-developed eukaryote cousins. But in any story about the early rise of oxygen on Earth, Kump and Barley remind us⁴, tracking the losses of early oxygen is as important as deciphering its sources. ■

Timothy W. Lyons is in the Department of Earth Sciences, University of California, Riverside, California 92521, USA.

e-mail: timothy.lyons@ucr.edu

1. Farquhar, J. & Wing, B. A. *Earth Planet. Sci. Lett.* **213**, 1–13 (2003).
2. Bekker, A. *et al. Nature* **427**, 117–120 (2004).
3. Brooks, J. J., Logan, G. A., Buick, R. & Summons, R. *Science* **285**, 1033–1036 (1999).
4. Kump, L. R. & Barley, M. E. *Nature* **448**, 1033–1036 (2007).
5. Pavlov, A. A. & Kasting, J. F. *Astrobiology* **2**, 27–41 (2002).
6. Holland, H. D. *Geochim. Cosmochim. Acta* **66**, 3811–3826 (2002).
7. Kasting, J. F., Egger, D. H. & Raeburn, S. P. *J. Geol.* **101**, 245–257 (1993).
8. Li, Z.-X. A. & Lee, C.-T. A. *Earth Planet. Sci. Lett.* **228**, 483–493 (2004).
9. Kopp, R. E., Kirschvink, J. L., Hillburn, I. A. & Nash, C. Z. *Proc. Natl Acad. Sci. USA* **102**, 11131–11136 (2005).
10. Bjerrum, C. J. & Canfield, D. E. *Geochim. Geophys. Geosyst.* **5**, doi:10.1029/2004GC000713 (2004).
11. Rosing, M. T. & Frei, R. *Earth Planet. Sci. Lett.* **217**, 237–244 (2004).
12. Anbar, A. D. *et al. Geochim. Cosmochim. Acta* **71** (supplement), A24 (2007).
13. Ohmoto, H., Watanabe, Y., Ikemi, H., Poulson, S. R. & Taylor, B. E. *Nature* **442**, 908–911 (2006).
14. Farquhar, J. *et al. Nature* (in the press).
15. Canfield, D. E. *Nature* **396**, 450–453 (1998).
16. Fike, D. A., Grotzinger, J. P., Pratt, L. M. & Summons, R. E. *Nature* **444**, 744–747 (2006).

MATERIALS SCIENCE

Polymer crystals downsized

Stephen Z. D. Cheng

A record has been broken — the smallest-ever polymer crystals have been prepared. Their surprising shape provides crucial evidence that might help explain how polymers crystallize.

How small can a crystal be and yet still be defined as a crystal? In other words, what is the smallest size at which a crystal exhibits its solid-state properties? At present, a quantitative, universal answer to this question does not exist, but Weber *et al.*¹ may be edging towards an answer, at least for polymer crystals. Reporting in *Nano Letters*, they describe the preparation of uniform-sized, single-crystal nanoparticles made from polyethylene — one of the cheapest and most widely used polymers. This casts light on fundamental aspects of polymer crystallization and on the technological issues involved in generating such crystals for specific applications.

Polymer scientists are keen to be able to

follow the crystallization of a single polymer chain. One approach would be to restrict the space available in which polymer molecules can crystallize, for instance by placing them in a nanotube. A nanotube would only accommodate a single, extended polymer chain; such a molecule would certainly not form a crystal. However, if the diameter of the nanotube was increased, enough space would become available to allow the polymer chain to fold, and so to form an ordered, three-dimensional arrangement of its atoms — a crystal. But what size would this lattice need to be for it to be called a crystal?

For now, the nanotube approach remains a thought experiment. Instead, researchers

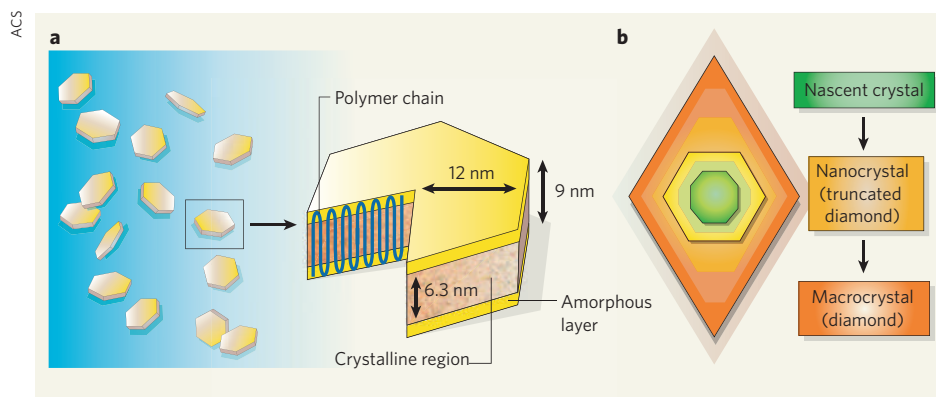


Figure 1 | Polyethylene crystal growth. **a**, Weber *et al.*¹ have prepared the smallest nanocrystals of polyethylene reported so far. The hexagonal structures comprise an inner crystalline layer sandwiched between two thin amorphous polymer sheets. The small size of the crystals is attributed to compact folding of the polymer chains. **b**, Large polyethylene crystals are diamond-shaped. One proposed explanation for how the hexagonal nanocrystals grow into diamonds is that two planes of the nanocrystals crystallize faster than the others. These planes therefore narrow as the crystal grows, and eventually disappear altogether. (Figure modified, with permission, from ref. 1.)

have tried to limit the size of polymer crystals by isolating as small a number of polymer molecules as possible. The most frequently used method is freeze-drying, in which very dilute solutions of polymers are quenched, which then enables individual molecules to be separated. These solutions are sprayed onto a substrate, where the solvent is rapidly evaporated to yield nanoparticles that contain low numbers of molecules; in the best cases, individual molecules are obtained. If the isolated polymer molecules are crystallizable, the nanoparticles can be converted into single crystals by heating them up to their melting point and then quenching them at a lower temperature. Such crystals are known as single-chain single crystals². But there are two problems with this approach: the crystals cannot be produced in a large quantity, and their size cannot be precisely controlled.

Weber *et al.*¹ now report the preparation of polyethylene nanocrystals, using a nickel-catalysed process that allows the polymer to be made in water. Under these conditions, nanoparticles of a uniform size form as soon as the polymerization reaction begins, and then quickly crystallize. The reaction is performed at 15 °C, which is more than 100 °C lower than the melting point of polyethylene; this is the lowest temperature at which anyone has been able to crystallize the polymer, including those using the freeze-drying experiments. The 'supercooling' involved results in a higher degree of chain folding than has previously been seen — in other words, the molecule crystallizes in a more confined space.

Weber and colleagues' report¹ provides fresh insight into polymer crystallization. The observation that single crystals, rather than crystal aggregates, form at low temperatures is of great interest as it provides clues about how crystal nuclei form and grow. Each nanocrystal comprises a hexagonal lamellar crystal sandwiched between two amorphous layers (folded surfaces), similar to the

structure of a hamburger (Fig. 1a). These are the smallest polyethylene crystals reported so far, with a width of around 25 nanometres and a thickness of 9 nm, including the amorphous layers (the crystalline thickness is 6.3 nm). Each crystal contains about 14 polymer chains. The faceted crystals clearly indicate a 'nucleation-limited' growth mechanism for these chain molecules, in which crystal growth follows the planes of the crystal lattice³.

Larger polyethylene crystals are known to have a diamond (lozenge) shape. If the hexagonal nanocrystals represent an early stage of crystal formation, how do they grow into a macroscopic diamond shape? Two pathways are proposed. One possibility is that two opposing planes of the hexagon — corresponding to the so-called {100} planes of the crystal lattice — crystallize more quickly than the others. Because the rest of the lattice grows more slowly, the edges of the hexagon corresponding to the {100} planes are forced to narrow as the crystal grows, until

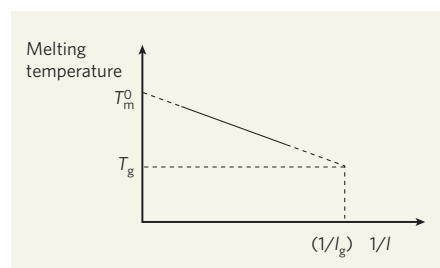


Figure 2 | Estimating minimum crystal sizes. Gibbs–Thomson graphs plot the melting temperature of a polymer against the inverse of the lamellar thickness (l) of the polymer's crystals. The line can be extrapolated (dashed sections) from experimental data to an infinite thickness ($1/l = 0$) to determine the melting points of polymers at thermodynamic equilibrium (T_m^0), a state that most polymers do not attain. Extrapolation to the glass transition temperature (T_g) — the temperature below which polymer molecules become immobile — corresponds to the smallest value of l (l_g) that can be obtained.

eventually the two planes are exhausted and disappear (Fig. 1b). Alternatively, it could be that the embryonic crystals adopt a hexagonal structure that grows quickly when the crystals are small, but take on the thermodynamically more stable diamond-shaped structure when they reach a critical size. Determining which pathway applies will lead to a new understanding of polymer crystallization.

So are Weber and colleagues' polymer crystals¹ the smallest that can be made? One way of judging this is to use the Gibbs–Thomson equation, which defines a linear relationship between the melting temperature of lamellar polymer crystals and the reciprocal of the lamellar thickness (Fig. 2). Graphs of this relationship are commonly extrapolated to determine the melting temperature of a polymer at thermodynamic equilibrium — polymers seldom reach this state, and so the equilibrium melting temperature cannot be found experimentally. Gibbs–Thomson graphs can also be extrapolated to a polymer's glass transition temperature — the point at which polymers turn from a rubbery to a glassy amorphous solid, and the stage at which all molecular mobility is arrested. This point corresponds to the smallest lamellar thickness that can practically be obtained.

Unfortunately, determining the melting temperatures of single-chain crystals is difficult, because they are metastable. Upon heating in a melting-point experiment, polymers can easily shift from one metastable state to another that has thicker lamellar polymer crystals⁴. Some polymer crystals that are far from equilibrium do this even if they are heated at rates exceeding several thousand degrees per second⁵. Extreme care must therefore be taken when determining data for a Gibbs–Thomson graph. Weber *et al.*¹ observe a thickening of their crystals owing to reorganization of the polymer after crystallization. These results suggest that many previously reported data for Gibbs–Thomson graphs must be re-examined.

The discovery of a method¹ to create polyethylene dispersions in water opens up potential applications for this material — and for related polyolefin polymers — that are currently restricted to other polymer classes. For example, ultrathin coating films of polyethylene could now be made. Given the enormous importance of latexes in polymer technology, Weber and colleagues' report could be the beginning of a revolution in polyolefin materials and their applications. ■

Stephen Z. D. Cheng is at the Maurice Morton Institute and Department of Polymer Science, University of Akron, Akron, Ohio 44325-3909, USA.

e-mail: scheng@uakron.edu

1. Weber, C. H. M. *et al.* *Nano Lett.* **7**, 2024–2029 (2007).
2. Bu, H. *et al.* *J. Polym. Sci. B* **29**, 139–152 (1991).
3. Cheng, S. Z. D. & Lotz, B. *Polymer* **46**, 8662–8681 (2005).
4. Keller, A. & Cheng, S. Z. D. *Polymer* **39**, 4461–4487 (1998).
5. Minakov, A. A., Mordvintsev, D. A. & Schick, C. *Faraday Discuss.* **128**, 261–270 (2005).

A functional circuit underlying male sexual behaviour in the female mouse brain

Tali Kimchi¹, Jennings Xu¹ & Catherine Dulac¹

In mice, pheromone detection is mediated by the vomeronasal organ and the main olfactory epithelium. Male mice that are deficient for *Trpc2*, an ion channel specifically expressed in VNO neurons and essential for VNO sensory transduction, are impaired in sex discrimination and male–male aggression. We report here that *Trpc2*^{−/−} female mice show a reduction in female-specific behaviour, including maternal aggression and lactating behaviour. Strikingly, mutant females display unique characteristics of male sexual and courtship behaviours such as mounting, pelvic thrust, solicitation, anogenital olfactory investigation, and emission of complex ultrasonic vocalizations towards male and female conspecific mice. The same behavioural phenotype is observed after VNO surgical removal in adult animals, and is not accompanied by disruption of the oestrous cycle and sex hormone levels. These findings suggest that VNO-mediated pheromone inputs act in wild-type females to repress male behaviour and activate female behaviours. Moreover, they imply that functional neuronal circuits underlying male-specific behaviours exist in the normal female mouse brain.

Males and females within a given animal species display identifiable differences in behaviours, mostly but not exclusively pertaining to sexual and social responses. Although these represent the most obvious examples of behavioural variability within a species, the basic principles underlying sexual dimorphism of brain function are largely unknown. Moreover, with few exceptions, the search for unique structures and circuits in male and female brains that parallel the dimorphism of peripheral sexual organs has so far met little success^{1–5}.

In many animals species- and sex-specific behaviours are orchestrated by pheromonal cues. Recent studies in rodents have uncovered the dual role of the main olfactory epithelium and the vomeronasal organ (VNO) in pheromones controlling mating, aggression and gender identification⁶. Genetic ablation of the TRPC2 channel, a signalling component essential to VNO function, leads to indiscriminate courtship and mounting behaviour of *Trpc2*^{−/−} male mice towards both males and females, suggesting an essential role of the vomeronasal system in sex identification^{7–10}. Furthermore, recent recording, genetic silencing and tracing experiments in the mouse have revealed the involvement of the main olfactory epithelium and associated central pathways in pheromone-mediated responses^{11–14}.

To study the role of the VNO in female sexual receptivity, we introduced a sexually experienced male to the home cage of either *Trpc2*^{+/+}, *Trpc2*^{+/-} or *Trpc2*^{−/−} females. As expected, oestrous *Trpc2*^{+/+} and *Trpc2*^{+/-} females were sexually receptive, allowing intensive olfactory investigation of the anogenital region by the male, leading to successful mating within minutes. However, in a striking role reversal, *Trpc2*^{−/−} females were observed intensively investigating the anogenital region of the intruder males and vigorously attempting to mount them, eliciting aggressive responses from the males.

Male-like behaviours of *Trpc2*^{−/−} females

Female–female and female–male mounting in rodents has been observed mainly in laboratory rats as part of dominance or sexual solicitation, respectively^{15–17}. The behaviour observed in *Trpc2*^{−/−}

females may thus represent either the exaggeration of normal female responses, or abnormal male-like displays. We monitored unique characteristics of male sexual and courtship behaviours in *Trpc2*^{+/+}, *Trpc2*^{+/-} and *Trpc2*^{−/−} male and female residents towards female and male intruders. To avoid the aggressive behaviour of wild-type males while controlling for the presence of pheromones, castrated or bulbectomized male swabbed with male urine were used as male intruders.

Male-like sexual display was investigated by scoring the number of animals mounting the intruder (Fig. 1a), the average duration of mounting (Fig. 1b) and the latency (time taken) to mount (Supplementary Fig. 1a) in a 15 min assay. Because mounting can include aspects of dominance, we also monitored occurrences of pelvic thrusts as a more stringent criteria of sexual behaviour (Fig. 1c). Results from all four tests demonstrate that *Trpc2*^{+/+} and *Trpc2*^{+/-} females very rarely displayed characteristics of male-like sexual behaviour towards female intruders, while the majority of *Trpc2*^{−/−} females (Supplementary Video 1), *Trpc2*^{+/-} and *Trpc2*^{−/−} males showed robust mounting, pelvic thrusts and a short latency to mount. Remarkably, the behaviour of *Trpc2*^{−/−} females towards other females was statistically indistinguishable from that of heterozygous and mutant males. Analysis of the response to male intruders showed that only *Trpc2*^{−/−} males and females (Supplementary Video 2) displayed significant levels of male-like sexual behaviour towards males, and that their behaviour is both statistically indistinguishable from each other, and from their response to female intruders.

We further assessed male-specific courtship behaviours. Male mice engage and solicit females by raising the female rear with their snout. Also, when interacting with females, adult males emit ultrasonic vocalizations at high (30–110 kHz) frequencies, while adult females produce only a limited range of ultrasounds during female–female social investigation^{18–22}. In addition, males perform intense olfactory investigation of the female rear, while females focus on the head and body.

The scoring of solicitation (Fig. 1d), ultrasound duration and complexity, latency to whistle, number of animals emitting ultrasounds (Fig. 1e, Fig. 2, Supplementary Fig. 1b–d), and olfactory

¹Department of Molecular and Cellular Biology, Howard Hughes Medical Institute, Harvard University, Cambridge, Massachusetts 02138, USA.

investigation of anogenital region (Supplementary Fig. 1e) further confirmed that the behaviour of *Trpc2*^{-/-} females towards females cannot be distinguished from that of *Trpc2*^{-/-} and *Trpc2*^{+/-} males, and is very different from that of *Trpc2*^{+/+} and *Trpc2*^{+/-} females. Moreover, the behaviour of *Trpc2*^{-/-} males and females towards male and female intruders was similar and, when intruders were presented simultaneously, *Trpc2*^{-/-} females, as previously shown with *Trpc2*^{-/-} males^{9,10}, display no preference for either sex, engaging indiscriminately in sexual behaviour with both male and female intruders with equal frequency ($n = 6$, data not shown).

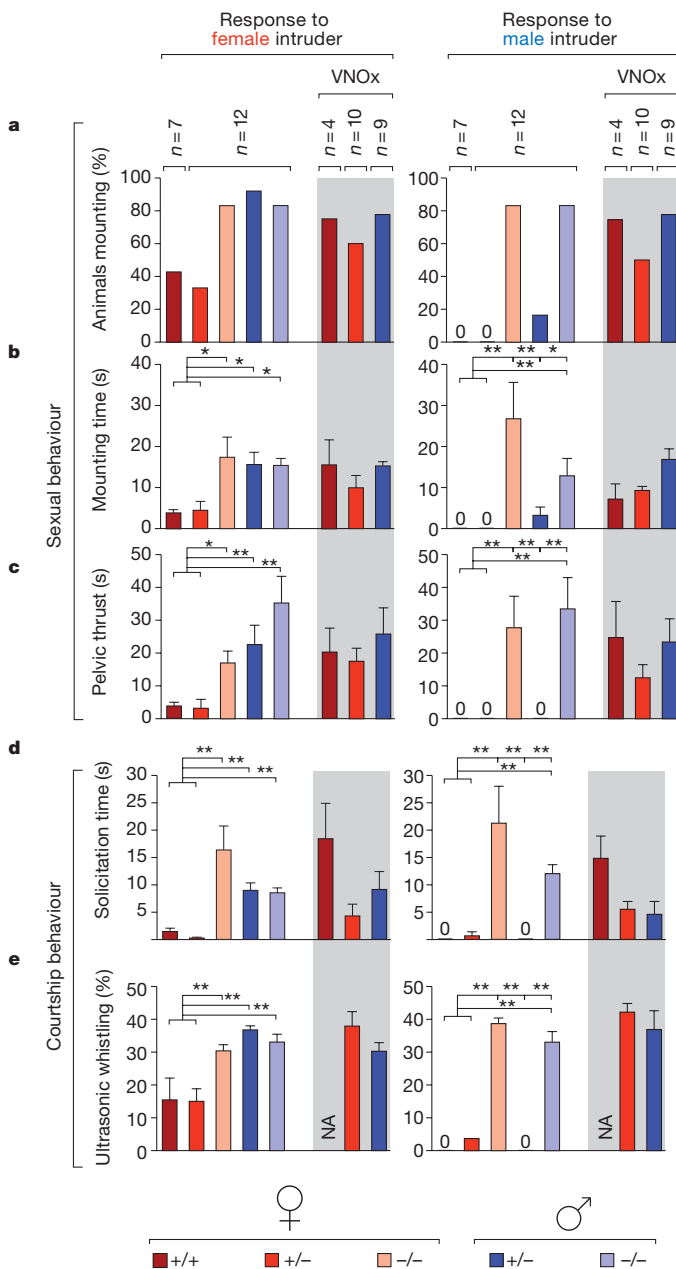


Figure 1 | Male-like sexual and courtship behaviours are displayed by *Trpc2*^{-/-} and VNOx females. Sexual (a–c) and courtship (d and e) behaviours towards female (left) and male (right) intruders observed in sexually naive adult *Trpc2*^{+/+}, *Trpc2*^{+/-} and *Trpc2*^{-/-} mice. VNOx mice had VNO surgical ablation. Responses to female intruders by *Trpc2*^{-/-} and VNOx females was indistinguishable from control and mutant males while responses to intruder males was similar to that of mutant males (error bars are s.e.m.; * $P < 0.05$; ** $P < 0.01$, one-way ANOVA followed by post-hoc Tukey Honest tests), suggesting a role of the VNO in sex discrimination in both males and females, and in repressing male-like behaviour in females. NA, not applicable.

Thus, the behaviour of *Trpc2*^{-/-} females highly resembles that typically exhibited by wild-type males interacting with females. Remarkably, in clear contrast with the normal, though rare occurrences of female mounting described in rodents^{15–17}, the mounting behaviour of the *Trpc2*^{-/-} females was not influenced by their oestrous stage (not shown), and rather than disappearing, it was maintained after sexual experience (Supplementary Fig. 2), and was not associated with dominance and aggression (see arena observation below). Moreover, some male-like behaviours shown by *Trpc2*^{-/-} females, such as pelvic thrusts and sexual solicitation with the snout (Supplementary Videos 1 and 2), are very rarely displayed by normal females. These observations strongly argue that the behaviour of *Trpc2*^{-/-} females does not simply result from a female hypersexual state, but represents genuine abnormal male-like traits that are quantitatively and qualitatively different from normal female behaviour.

Surgical ablation of the VNO in adults

Because the *Trpc2* mutation eliminates VNO function throughout life, we reasoned that the male-like sexual behaviour of *Trpc2*^{-/-} females may result from the abnormal development of female behaviour circuits. Alternatively, it may reveal the de-repression of an existing male behaviour circuit that is normally masked in the female brain by inhibitory vomeronasal inputs. To distinguish between these hypotheses, we compared the behaviour of mice in which the VNO has been surgically removed in the adult to the behaviour of genetically deficient animals of the same age.

Adult olfactory marker protein (OMP)-*TLZ*^{+/+} males and females, in which olfactory expression of tau-LacZ permits direct visualization of main olfactory epithelium and VNO projections²³, had their VNO removed (VNOx) and their airways cleared daily for a week after surgery. Behavioural tests were performed three weeks after surgery, and animals were killed to assess the extent of VNO removal and accessibility of nasal airways (Fig. 3). These additional steps were made necessary by our observation that VNO ablation may cause bleeding and obstruction of the nasal cavity (Supplementary Fig. 3), resulting in behaviour patterns similar to that of olfactory-deficient mice, such as severe deficits in sexual and aggressive behaviour. We also performed VNO surgical ablation in *Trpc2*^{-/-} mice, and showed that mutant animals with or without intact VNOs of a given gender ($n = 9$ each gender) exhibited identical behaviours (not shown), ensuring that no remnants of VNO function can be identified in *Trpc2*^{-/-} animals.

Our data showed that the sexual and courtship behaviours of VNOx males towards male and female intruders are indistinguishable from those of *Trpc2*^{-/-} mutants with intact VNOs. (Fig. 1, Fig. 2 and Supplementary Fig. 1, right side of each histogram). Further, *Trpc2*^{+/-} and *Trpc2*^{-/-} VNOx females exhibited most male-like traits at levels similar to that of *Trpc2*^{-/-} females. These included mounting, pelvic thrust, latency to mount, ultrasound vocalization and olfactory investigation (Supplementary Video 3). We obtained similar results when the VNO removal was performed on C57BL/6J adults (Supplementary Fig. 4), in direct contrast with previously published studies^{24,25}. Because these earlier studies did not control for the possible occlusion of the nasal cavity (a frequent occurrence after standard VNO surgical removal), it is possible that additional olfactory deficits have confounded the interpretation of the data.

Thus, the loss of VNO function in adulthood resulted in altered sexual behaviour in males and in sudden sex-reversal of female behaviour, demonstrating the requirement for sustained control by VNO inputs to ensure normal sex discrimination in males and females, and female-specific sexual behaviour. In addition, slight differences in the behaviour of *Trpc2*^{-/-} and VNOx females suggest that VNO activity plays a minor role during development.

Behaviour under semi-natural conditions

When studying complex and dynamic behaviours, such as social interactions, confined experimental conditions could cause the

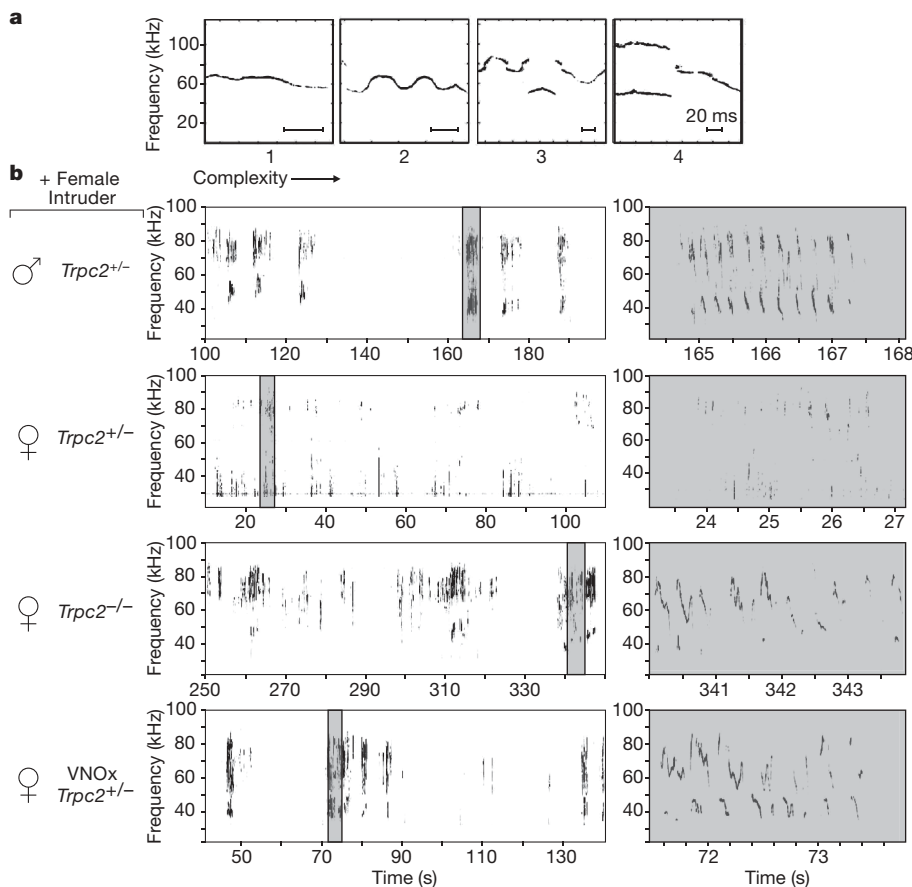
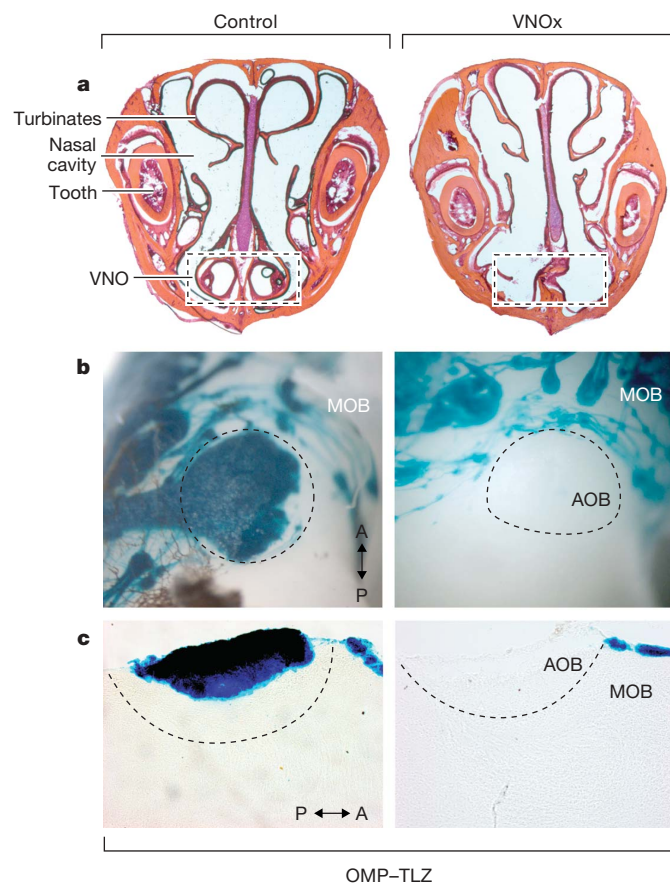


Figure 2 | Ultrasonic vocalization by male and female mice in resident-intruder assays.

a, Ultrasonic complexity index. Whistle clusters with no pitch jump and temporal overlap in frequency received a low complexity score (1–2), while clusters of whistles with pitch jump and extensive temporal overlap received a high complexity score (3–4). **b**, Representative examples of ultrasonic vocalizations emitted by a *Trpc2*^{+/–} male and female, a *Trpc2*^{–/–} female and a VNOx *Trpc2*^{+/–} female in the presence of a female intruder. Except for the *Trpc2*^{+/–} female, other examples show whistle clusters with highest complexity score (3–4). Power below 35 kHz was truncated.



animal to present a limited and even impaired behavioural repertoire. Therefore, we also decided to test the role of VNO-mediated pheromone detection in female mice under semi-natural conditions.

Groups of four *Trpc2*^{+/–} and *Trpc2*^{–/–} females were first kept separated in each half of a large enclosure with enriched environment, and were scored for social behaviours. In contrast to the *Trpc2*^{+/–} females, the *Trpc2*^{–/–} females exhibited high levels of social interactions such as mounting, anogenital olfactory investigation, as well as defensive behaviour resulting from mounting attempts by other females (Fig. 4, stage 1). Importantly, neither the *Trpc2*^{–/–} nor the *Trpc2*^{+/–} group established any recognizable dominant–subordinate social hierarchy.

Sexually experienced wild-type males were then introduced into each side of the enclosure. Remarkably, unlike *Trpc2*^{+/–} females (Supplementary Video 4), *Trpc2*^{–/–} females (Supplementary Video 5) intensely chased the males in attempt to investigate their anogenital region and to mount (Fig. 4, stage 2). Again, we found no dominant–subordinate social hierarchy established in either group, and no correlation between the level of female–male mounting in *Trpc2*^{–/–} individuals and the level of aggression towards each other or towards males. Thus, in contrast with normally occurring female–female mounting, the mounting behaviour of *Trpc2*^{–/–} females is unrelated to dominance. Moreover, as shown above, it is not a sexual solicitation

Figure 3 | Surgery leads to a complete removal of the vomeronasal organ (VNOx) while the nasal airways stay clear. **a**, Coronal sections (50 μm) through the anterior part of the skull of control (left) and VNOx mice (right) were stained with haematoxylin and eosin, showing full removal of the bilateral VNO structure while the nasal airway is kept open. **b**, **c**, X-Gal staining of the olfactory bulb of OMP-ires-tauLacZ (OMP-TLZ) control (left) and VNOx mice. **b**, Whole-mount staining. Dorsal view. **c**, Parasagittal sections (50 μm) served as controls for the complete disappearance of all VNO projections to the accessory olfactory bulb. A, anterior; P, posterior; AOB, accessory olfactory bulb; MOB, main olfactory bulb.

of non-responsive males by oestrous females as described in rats^{15,16}, and instead has characteristics of genuine male sexual behaviour.

Most females within the arena became pregnant and 15 to 22 pups were delivered per group. The successful mating of males with *Trpc2*^{-/-} females was probably due to increased aggressive behaviour and mating attempts of the males with the *Trpc2*^{-/-} females. We subsequently tested maternal aggression of lactating females towards intruder males by adding a strange male from the CD1 strain to the enclosure. All lactating *Trpc2*^{+/-} females attacked the intruder male and showed low sexual receptivity. In contrast, the intruder male evoked a low level of aggression from the *Trpc2*^{-/-} females, which appeared highly sexually receptive (Fig. 4, stage 3). Our findings confirm previous results showing low aggression from lactating *Trpc2*^{-/-} females¹⁰ while contradicting the described deficiency in sexual receptivity of VNOx females²⁴.

Finally, we investigated female maternal and lactating behaviours. Both *Trpc2*^{+/-} and *Trpc2*^{-/-} females kept their litters in one common breeding nest where the females nursed the pups in turn. In the first and second days after birth, both groups spent a similar amount of time with their pups. However, during the following days, *Trpc2*^{-/-} females exhibited a significant decrease in time spent in the breeding nest, while *Trpc2*^{+/-} females spent most of their time in the breeding nest. On the last day (day 14), the partition between the

Trpc2^{-/-} and *Trpc2*^{+/-} females was lifted to allow both groups to interact. Strikingly, the *Trpc2*^{-/-} females immediately abandoned the breeding nest to explore the *Trpc2*^{+/-} territory, while *Trpc2*^{+/-} females remained in breeding nests and continued to nurse their pups (Fig. 4, stage 4). These findings indicate that *Trpc2*^{-/-} females display a deficiency in maternal behaviour.

Discussion

We have shown here that *Trpc2*^{-/-} females exhibit robust male-like sexual and courtship behaviours and display a reduction in female-specific behaviours such as maternal aggression and nesting. These findings suggest that VNO-mediated inputs repress male-like sexual and courtship behaviours in females, and activate and sustain female maternal behaviours.

Previous studies have suggested a central role for sex hormones in the regulation of male- and female-specific behaviours²⁶. We found that body weight and oestrous cycles of *Trpc2*^{-/-} females (Fig. 5a) appeared normal. Moreover, radioimmunoassay for testosterone and 17 β -estradiol in the serum of *Trpc2*^{-/-} and *Trpc2*^{+/-} males and females reveal that sex steroid levels remain within the normal range for each gender (Fig. 5a). A minor increase in free testosterone level was observed in *Trpc2*^{-/-} females that still vastly differs from typical male levels and from the amounts of exogenous testosterone

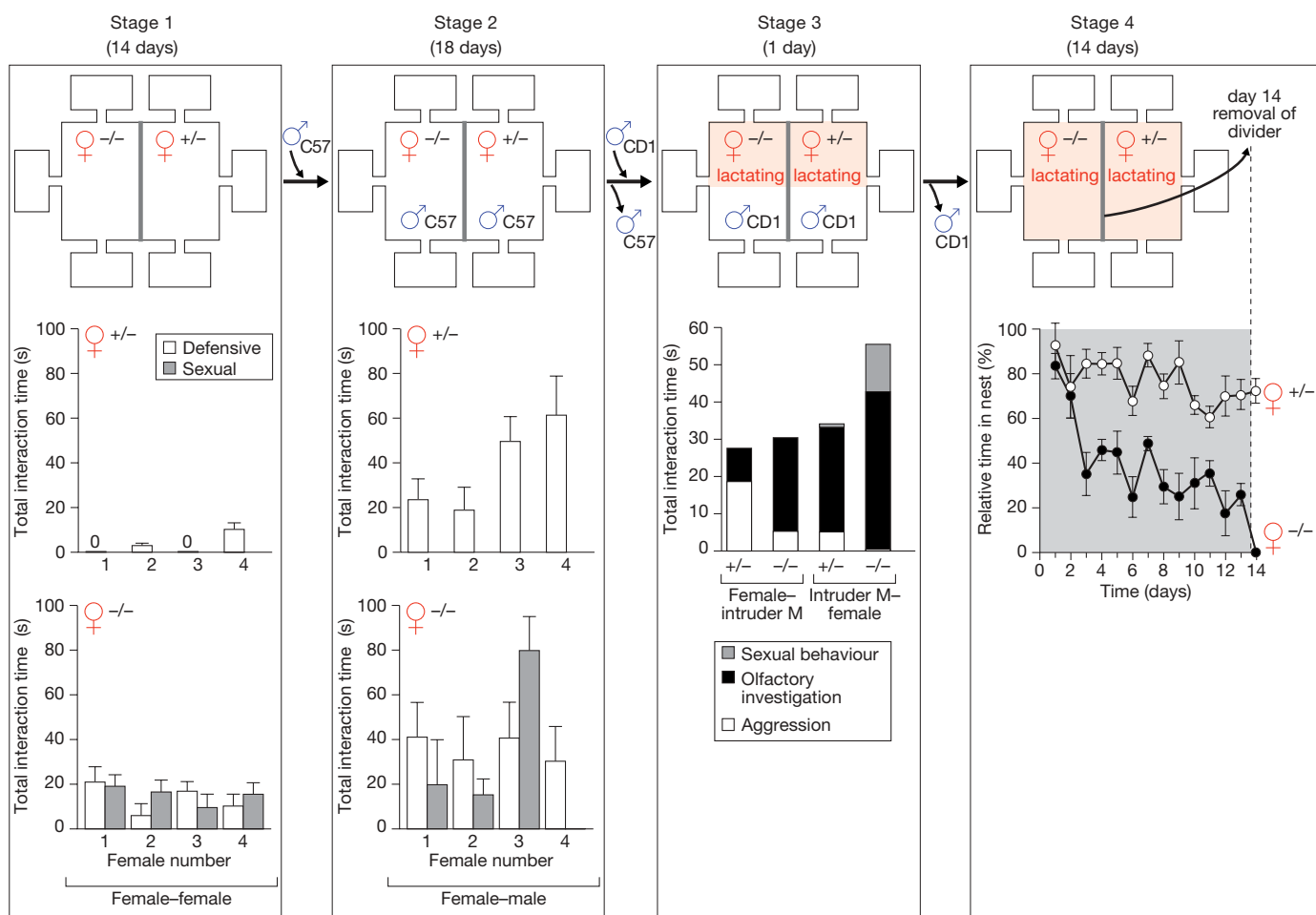


Figure 4 | Social behaviour of *Trpc2*^{+/-} and *Trpc2*^{-/-} females in semi-natural conditions. Stages 1 and 2 show duration of defensive (white) and sexual (grey) behaviours of resident females with each other (stage 1) and towards a C57BL/6J adult male (stage 2). Stage 3 shows duration of sexual behaviour (grey), olfactory investigation (black) and aggression (white) of the resident lactating females towards an intruder CD1 adult male (left two columns) and behaviour of intruder CD1 towards the resident females (right

two columns). Stage 4 shows relative time spent by lactating *Trpc2*^{+/-} and *Trpc2*^{-/-} females in the breeding nest before and after (day 14) the divider was lifted. In the enriched environment *Trpc2*^{-/-} females exhibit high levels of male-like sexual behaviour (stages 1 to 2). They exhibit no apparent deficiency in sexual receptivity but show reduction in maternal (stage 3) and lactating behaviour (stage 4). Stages 1–4, error bars are s.e.m.

required to affect normal female behaviour^{27,28}. Thus, hormonal changes do not seem to underlie the expression of male-like behaviours in *Trpc2*^{-/-} females.

The prevailing model for the sexual dimorphism of behaviours is that the sex hormone testosterone initiates the development of male-specific circuitry in the central nervous system and the activation of male-specific neuronal networks in adulthood²⁶. However, our results clearly reveal that a functional neuronal network mediating male sexual behaviour develops and persists in females. These findings suggest a new model of sexual dimorphism in which the effector circuits of both male and female behaviours exist in the brain of each sex, and are switched on or off by gender-specific sensory modulators (Fig. 5b). Interestingly, one or a few classes of olfactory receptors have been shown in *Drosophila* to mediate the specificity of male and female sexual responses in a similar way^{29,30}. Further support for our model in the mouse will involve the identification of specific receptor inputs that inhibit the expression of male responses in the female brain. Moreover, a similar inhibitory control of female behaviour may yet be found in the male brain.

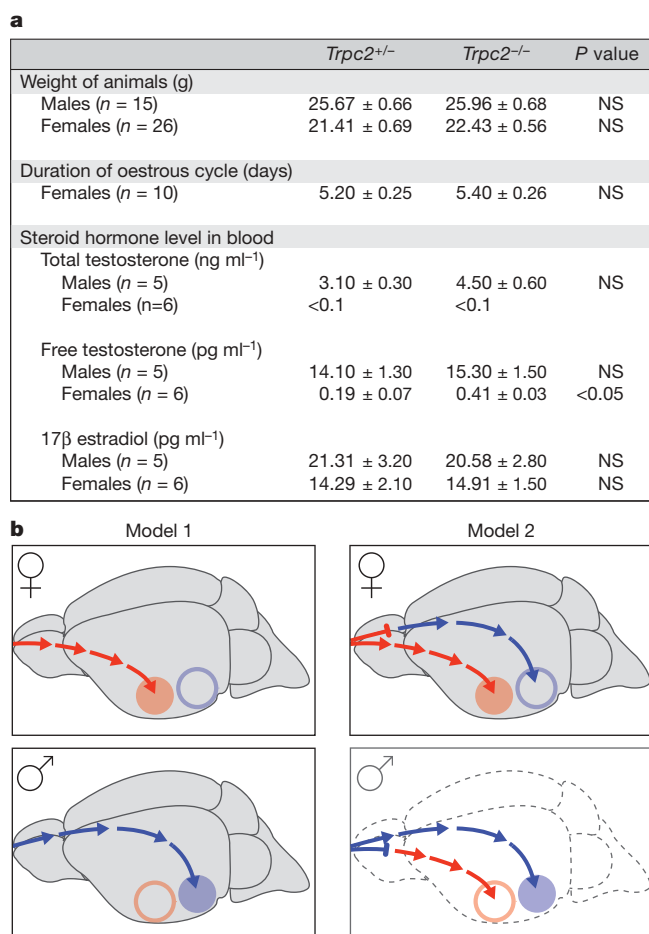


Figure 5 | Sexual dimorphic traits in *Trpc2*^{-/-} mutants and regulation of sexually dimorphic behaviour. **a, *Trpc2*^{-/-} female body weight, oestrous cycle and steroid hormone levels are in the range of that of heterozygote females. Values are means ± s.e.m. NS = $P > 0.1$, Student's *t*-tests for independent samples. **b**, Prevailing views on neuronal networks underlying sexual dimorphic behaviours (model 1) have assumed the existence of sex-specific circuits throughout the brain, including male (blue) and female (red) specific sensory and effector networks. In contrast, data resulting from VNO deficiency in females support a different model (model 2), according to which effector networks for sex-specific behaviours are present in both females (our data) and males (dotted lines, to be determined), while the sensory switch is sexually dimorphic.**

METHODS SUMMARY

Behavioural assays. We used sexually naive, 2–3-month-old *Trpc2*^{+/+}, *Trpc2*^{+/-}, *Trpc2*^{-/-} (ref. 9), and C57BL/6J × 129/Sv mice of mixed genetic background as well as C57BL/6J mice in the behaviour experiments. Mice were tested in 15 min intruder–resident assays in which individually housed residents mice were exposed to sexually naive female, and either castrated or olfactory bulbectomized male C57BL/6J intruders. Ultrasound vocalizations by the resident mice were recorded during intruder–resident assays. Stored recordings were processed using a custom MATLAB-based program as described¹⁸. Social behaviour within a colony of *Trpc2*^{+/+} and *Trpc2*^{-/-} mice was recorded under semi-natural conditions within a large (120 cm × 90 cm × 80 cm), environmentally enriched enclosure. The behaviour of the mice was recorded by low-light-sensitive video cameras that were connected to a custom-designed PC-based recording unit (Protech PC). The social behaviour (courtship and sexual behaviours) in both the intruder–resident assays and in the semi-natural conditions were scored in the recorded videos using Observer Video Pro software (Noldus). **Surgical VNO removal.** The VNO was surgically ablated from adult, 8–10-week-old, sexually naive mice (*Trpc2*^{+/+} × OMP-tauLacZ and *Trpc2*^{-/-} × OMP-tauLacZ). For the first week after surgery, VNOx mice were anaesthetized daily, their nostrils were rinsed with 0.9% saline solution, and any blood clots were gently aspirated. VNOx mice were allowed at least three weeks to recover before behavioural testing.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 15 May; accepted 17 July 2007.

Published online 5 August 2007.

1. Simerly, R. B. Wired for reproduction: organization and development of sexually dimorphic circuits in the mammalian forebrain. *Annu. Rev. Neurosci.* **25**, 507–536 (2002).
2. Sagovia, S. & Guillemon, A. Sexual dimorphism in the vomeronasal pathway and sex differences in reproductive behaviors. *Brain Res. Brain Res. Rev.* **18**, 51–74 (1993).
3. Shah, N. M. *et al.* Visualizing sexual dimorphism in the brain. *Neuron* **43**, 313–319 (2004).
4. Manoli, D. S. *et al.* Male-specific fruitless specifies the neural substrates of *Drosophila* courtship behaviour. *Nature* **436**, 395–400 (2005).
5. Stockinger, P., Kvitsiani, D., Rotkopf, S., Tirián, L. & Dickson, B. J. Neuronal circuitry that governs *Drosophila* male courtship behavior. *Cell* **121**, 795–807 (2005).
6. Dulac, C. & Wagner, S. Genetic analysis of brain circuits underlying pheromone signaling. *Annu. Rev. Genet.* **40**, 449–467 (2006).
7. Liman, E. R., Corey, D. P. & Dulac, C. TRP2: A candidate transduction channel for mammalian pheromone sensory signaling. *Proc. Natl Acad. Sci. USA* **96**, 5791–5796 (1999).
8. Lucas, P., Ukhonov, K., Leinders-Zufall, T. & Zufall, F. A diacylglycerol-gated cation channel in vomeronasal neuron dendrites is impaired in TRPC2 mutant mice. *Neuron* **40**, 551–561 (2003).
9. Stowers, L., Holy, T. E., Meister, M., Dulac, C. & Koentges, G. Loss of sex discrimination and male-male aggression in mice deficient for TRP2. *Science* **295**, 1493–1500 (2002).
10. Leypold, B. G. *et al.* Altered sexual and social behaviors in *trp2* mutant mice. *Proc. Natl Acad. Sci. USA* **99**, 6376–6381 (2002).
11. Yoon, H., Enquist, L. W. & Dulac, C. Olfactory inputs to hypothalamic neurons controlling reproduction and fertility. *Cell* **123**, 669–682 (2005).
12. Boehm, U., Zou, Z. & Buck, L. B. Feedback loops link odor and pheromone signaling with reproduction. *Cell* **123**, 683–695 (2005).
13. Lin, D. Y., Zhang, S. Z., Block, E. & Katz, L. C. Encoding social signals in the mouse main olfactory bulb. *Nature* **434**, 470–477 (2005).
14. Mandiyan, V. S., Coats, J. K. & Shah, N. M. Deficits in sexual and aggressive behaviors in *Cnga2* mutant mice. *Nature Neurosci.* **8**, 1660–1662 (2005).
15. Fang, J. & Clemens, L. Contextual determinates of female-female mounting in laboratory rats. *Anim. Behav.* **57**, 545–555 (1999).
16. Afonso, V. M. & Pfaus, J. G. Female mounting of male rats: the role of sexual experience and hormone priming. *Horm. Behav.* **49**, 30–37 (2006).
17. Beach, F. A. Factors involved in the control of mounting behavior by female mammals. In *Perspectives in Reproduction and Sexual Behavior: a Memorial to William C. Young* (ed. Diamond, E.) 88–131 (Indiana Univ. Press, Bloomington, 1968).
18. Holy, T. E. & Guo, Z. Ultrasonic songs of male mice. *PLoS Biol.* **3**, 2177–2186 (2005).
19. White, N. R., Prasad, M., Barfield, R. J. & Nyby, J. 40 and 70-kHz vocalization of mice, *Mus musculus*, during copulation. *Physiol. Behav.* **63**, 467–473 (1998).
20. Maggio, J. & Whitney, G. Ultrasonic vocalizing by adult female mice (*Mus musculus*). *J. Comp. Psychol.* **99**, 420–436 (1985).
21. Gourbal, B. E. F., Barthelemy, M., Petit, G. & Gabrion, C. Spectrographic analysis of the ultrasonic vocalizations of adult male and female BALB/c mice. *Naturwissenschaften* **91**, 381–385 (2004).
22. Moles, A. & D'Amato, F. Ultrasonic vocalization by female mice in the presence of a conspecific carrying food cues. *Anim. Behav.* **60**, 689–694 (2000).

23. Mombaerts, P. *et al.* Visualizing an olfactory sensory map. *Cell* **87**, 675–686 (1996).
 24. Keller, M., Pierman, S., Douhard, Q., Baum, M. & Backer, J. The vomeronasal organ is required for the expression of lordosis behavior, but not sex discrimination in female mice. *Eur. J. Neurosci.* **23**, 521–530 (2006).
 25. Pankevich, D. E., Baum, M. J. & Cherry, J. A. Olfactory sex discrimination persists, whereas the preference for urinary odorants from estrous females disappears in male mice after vomeronasal organ removal. *J. Neurosci.* **24**, 9451–9457 (2004).
 26. Morris, J. A., Jordan, C. L. & Breedlove, S. M. Sexual differentiation of the vertebrate nervous system. *Nature Neurosci.* **7**, 1034–1039 (2004).
 27. Edwards, D. A. & Katherine, G. B. Early androgen treatment and male and female sexual behavior in mice. *Horm. Behav.* **2**, 49–58 (1971).
 28. Manning, A. Neonatal androgen and sexual behavior in female house mice. *Horm. Behav.* **5**, 19–31 (1974).
 29. Kurtovic, A., Widmer, A. & Dickson, B. J. A single class of olfactory neurons mediates behavioural responses to a *Drosophila* sex pheromone. *Nature* **446**, 542–546 (2007).
 30. Van der Goes van Naters, W. & Carlson, J. R. Receptors and neurons for fly odors in *Drosophila*. *Curr. Biol.* **17**, 606–612 (2007).
- Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.
- Acknowledgements** We thank R. Hellmiss for artistic work and illustrations, S. Sullivan for assistance with behavioural analysis, A. E. Launjuin and S. Yao for help with histochemistry, the Dulac laboratory for discussions and comments on the manuscript, T. E. Holy, B. Ölveczky and A. Kampf for help with the recording of ultrasonic vocalizations. The work was supported by the Howard Hughes Medical Institute (to C.D.), the NIH (to C.D.) and the Human Frontier Science Program (to T.K.).
- Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.D. (dulac@fas.harvard.edu) or T.K. (kimhi@fas.harvard.edu).

METHODS

Intruder–resident assay. *Trpc2*^{+/-} and *Trpc2*^{-/-} (ref. 9) sexually naive, 2–3-month-old, female and male littermates of the C57BL/6J × 129/Sv mixed genetic background were housed individually in a cage for one to two weeks before the beginning of the experiment. Experiments started at the beginning of the dark phase and were performed under dim red light.

The following two kinds of intruders (C57BL/6J) were introduced to the resident mouse cage: sexually naive, receptive female (as determined by vaginal smear), 7–8 weeks old; and sexually naive, castrated or olfactory bulbectomized male, 6–7 weeks old, swabbed with urine from intact wild-type males. Each 15 min assay was videotaped and scored for the following three behaviours: sexual behaviour (mounting, and mounting with pelvic thrust); solicitation behaviour (resident animal lifts intruder's rear with its snout); and olfactory investigation.

Recording and processing of ultrasonic vocalization. Sounds over the frequency range of 20 Hz–110 kHz were recorded with a microphone and amplifier (Bruel & Kjaer) and digitized at 250 kHz, 16 bits (National Instruments) and saved to disk within a custom-designed MATLAB-based program.

In brief, the MATLAB-based program converted the stored waveforms to sonograms (512 sample/blockform, time resolution of 1.02 ms and a frequency of 0.98 kHz), removing white noise outside the range 25–110 kHz; it identified and presented each ultrasonic mouse whistle. The MATLAB-based program was used to estimate the whistling activity level by summing the overall whistling time out of the total whistling clusters time for each observation. Whistle clusters were classified by the criteria that there existed at least five whistles where the period of time between each whistle was less than 500 ms, and additionally, that the beginning and end whistle were separated from the previous and next whistles, respectively, by at least 500 ms.

To confirm that the resident mouse was the source of the ultrasonic vocalizations we recorded from assays in which either the resident or the intruder mouse was anaesthetized. We were only able to record robust ultrasonic vocalizations resembling those we recorded during the resident–intruder assays if the intruder mouse was anaesthetized and not the resident.

Semi-natural experimental set-up and procedure. Sexually naive, 3–4-month-old females ($n = 4$ each genotype) *Trpc2*^{-/-} and *Trpc2*^{+/-} mice, derived from four different litters, were used. The female mice were housed individually in a cage two weeks before the beginning of the test. To enable us to individually recognize the animals, before the beginning of the experiment the animals were anaesthetized (with 120 mg of ketamine per kg of mouse body weight and

10 mg kg⁻¹ xylazine) and marked with commercial hair dye. The mice were re-marked every two to three weeks throughout the experiment. The experiment was conducted twice with different animals.

Semi-natural enclosure set-up. The enclosure consisted of a large central arena constructed from transparent polycarbonate boards that were connected to six peripheral standard mouse cages by short transparent tubes. A removable oblique polycarbonate board (90 × 80 cm) served as a divider. The floor of the arena and cages were spread with bedding, scattered strips of towel paper, cotton pieces, and shelter boxes. In addition, each half of the arena contained a central platform with rodent pellets and water freely supplied.

The enclosure was placed in a temperature- and light-controlled (12 h:12 h light/dark cycle) room furnished with infrared lights and equipped with low-light-sensitive cameras mounted above and around the enclosure. All cameras were connected to a digital video recording unit (ProtechPC).

The behaviour was recorded daily for 10–20 min every hour, from the beginning of the dark phase until 2 h before the light phase. The data were scored using the Observer Video Pro software (Noldus).

Testosterone and oestrogen blood level measurements. Animals were killed by overdose of 2.5% avertin. Blood was removed from the heart, and serum or plasma was separated and stored at -20 °C until analysed. Concentrations of testosterone and 17 β -estradiol were measured using a radioimmunoassay RIA Kit (MP Biomedicals).

Surgical VNO removal. Mice were anaesthetized (120 mg kg⁻¹ ketamine and 10 mg kg⁻¹ xylazine), placed supine in a head holder, and the lower jaw was gently opened. A midline incision was made in the soft palate, extending rostrally from behind the first palatal ridge to the incisors, and the underlying bone was exposed. The caudal end of the vomer bone was cut and the VNO was removed bilaterally. Low-pressure vacuum was used to clear blood from the mouse and nostrils during the surgery. The VNO cavity was packed with absorbable gel foam (Pharmacia) and the incision was closed with veterinary sterile tissue adhesive (Tissumend II).

After testing, VNOx mice were killed to confirm complete VNO removal. To confirm complete degeneration of axonal projections to the accessory olfactory bulb, we performed whole-mount X-Gal staining of the olfactory bulbs, as previously described²³, followed by 50 μ m parasagittal sections of the stained tissue, for more detailed examination. To confirm that there were no blood clots blocking the olfactory airways, skulls of VNOx mice were decalcified in 10% EDTA (pH = 7.4), sectioned coronally (50 μ m) on a cryostat, counterstained with haematoxylin and eosin, and dehydrated.

IGF and FGF cooperatively establish the regulatory stem cell niche of pluripotent human cells *in vitro*

Sean C. Bendall^{1,2*}, Morag H. Stewart^{1*}, Pablo Menendez^{1†}, Dustin George², Kausalia Vijayaragavan¹, Tamra Werbowetski-Ogilvie¹, Veronica Ramos-Mejia¹, Anne Rouleau¹, Jiabi Yang¹, Marc Bossé¹, Gilles Lajoie² & Mickie Bhatia¹

Distinctive properties of stem cells are not autonomously achieved, and recent evidence points to a level of external control from the microenvironment. Here, we demonstrate that self-renewal and pluripotent properties of human embryonic stem (ES) cells depend on a dynamic interplay between human ES cells and autologously derived human ES cell fibroblast-like cells (hdFs). Human ES cells and hdFs are uniquely defined by insulin-like growth factor (IGF)- and fibroblast growth factor (FGF)-dependence. IGF 1 receptor (IGF1R) expression was exclusive to the human ES cells, whereas FGF receptor 1 (FGFR1) expression was restricted to surrounding hdFs. Blocking the IGF-II/IGF1R pathway reduced survival and clonogenicity of human ES cells, whereas inhibition of the FGF pathway indirectly caused differentiation. IGF-II is expressed by hdFs in response to FGF, and alone was sufficient in maintaining human ES cell cultures. Our study demonstrates a direct role of the IGF-II/IGF1R axis on human ES cell physiology and establishes that hdFs produced by human ES cells themselves define the stem cell niche of pluripotent human stem cells.

Self-renewal and multi-lineage developmental potential define the unique properties attributed to stem cells. *In vivo*, these properties are not autonomous to stem cells, and recent evidence points to a level of external control from the microenvironment that defines the stem cell niche¹. Seemingly, human ES cells represent an exception to this paradigm, where human ES cells can be isolated from their blastocytic microenvironment and propagated in *in vitro* culture systems^{2–5}. Human ES cells are defined by both robust self-renewal capacity and maintenance of pluripotent developmental capacity *in vitro* and *in vivo*^{2,6}. Human ES cell lines have been derived and cultured in several systems, including mouse embryonic fibroblast-conditioned media (MEF-CM) and in the absence of non-human proteins for the purpose of future clinical transplantation^{2–5}. Despite the evolution of these culture conditions and the addition of several factors, all require supplementation with basic fibroblast growth factor (bFGF) to sustain human ES cell potential. Unlike leukaemia inhibitory factor (LIF) or bone morphogenetic protein (BMP) used in the maintenance of murine ES cells⁷, the basis for using bFGF, or a comprehensive definition of factors directly acting on human ES cells, remains unknown. Using a proteomic screening approach, we reveal that human-ES-cell-derived fibroblast-like cells (hdFs) act as a supportive niche via IGF-II production, and that IGF-II has a direct role in the survival and self-renewal of pluripotent human ES cells.

Deconstruction of human ES cell cultures via receptor expression

To understand better the factors capable of regulating human ES cells, we performed a proteomic screen to identify candidate human ES cell supportive factors present in MEF-CM. The most prevalent family of growth factors was insulin-like growth factor (IGF), of

which IGF-II was the best-represented (Supplementary Table 1). IGF-II is known to have a central regulatory role in mammalian pre-implantation and embryonic development^{8–10}. Notably, the receptor for IGF-II, IGF1R, was also identified in conjunction with the receptor for bFGF in a genomic analysis of human ES cells¹¹. As the IGF pathway had not been investigated in the context of human ES cells, we elected to characterize its role on the functional properties that define human ES cells.

We initially observed that human ES cells cultured in the absence of feeder layers can be divided into two morphologically distinct cellular compartments: Oct4 (also called POU5F1)-expressing colonies and surrounding non-colony cells. Colonies are comprised of undifferentiated human ES cells, whereas the surrounding cells are generated directly from the human ES cell colonies in both feeder and feeder-free cultures (Fig. 1a, b). These surrounding hdFs⁶ were observed in four independent human ES cell lines (Supplementary Fig. 1). On the basis of the requirement of bFGF in the maintenance of all human ES cell cultures reported so far^{12,13}, human ES cells would be predicted to express FGF receptors. Quantitative real-time polymerase chain reaction (qPCR) comparison of human ES cell cultures to non-colony hdF cultures demonstrated that, whereas *IGF1R* and *FGFR1* expression was equivalent in bulk human ES cell cultures, hdF cultures predominantly expressed *FGFR1*, but lacked expression of the human ES cell marker *Oct4* (ref. 6; see also Fig. 1c). As hdF cultures are derived directly from *Oct4*-expressing human ES cells⁶, this indicated that *FGFR1* and *IGF1R* were expressed on distinct populations within human ES cell cultures.

To examine directly the relevance of *IGF1R* expression relative to colony-forming human ES cells, IGF1R⁺ and IGF1R[−] cells were

¹McMaster Stem Cell and Cancer Research Institute, Michael G. DeGroote School of Medicine, and Department of Biochemistry, McMaster University, Hamilton, Ontario L8N 3Z5, Canada. ²Don Rix Protein Identification Facility, Department of Biochemistry, Schulich School of Medicine and Dentistry, University of Western Ontario, London, Ontario N6A 5C1, Canada. [†]Present address: Spanish Stem Cell Bank-Andalucian Branch, Instituto de Investigaciones Biomedicas, Granada 18100, Spain.

*These authors contributed equally to this work.

prospectively isolated from human ES cell cultures by fluorescence-activated cell sorting (FACS) (Fig. 1d). qPCR analysis revealed that the IGF1R⁺ cell fraction was highly enriched for *IGF1R* and *POU5F1* (*Oct4*) transcript, whereas *FGFR1* expression was surprisingly restricted to the IGF1R⁻ cells (Fig. 1e). Quantitative clonogenic analysis⁶ of human ES cell colony-initiating cells (hESC-CIC) revealed that the capacity to produce human ES cell colonies was almost exclusively found within the IGF1R⁺ fraction alone (Fig. 1f, g), providing direct functional evidence for human ES cells expressing *IGF1R*.

Although *FGFR1* represents the best candidate receptor for bFGF, *FGFR1* is only one of four putative human bFGF receptors. Similar to previous observations using unfractionated human ES cell cultures^{11,12,14,15}, *FGFR1* was the most highly expressed receptor (Fig. 1h). Upon isolation and comparison of IGF1R⁺ cells, IGF1R⁻ cells and hdFs to bulk cultures, we observed that all known FGF receptors are expressed on sub-fractions of the human ES cell culture that lack human ES cell CIC capacity or *Oct4* expression (Fig. 1i). Similarly, expression of *FGFR3* and *FGFR4* was not detectable by western blot or immunocytochemistry, whereas *FGFR2* could be detected at low levels in bulk human ES cell cultures by western blot, and immunocytochemistry analysis indicated that expression originates from rare cells surrounding the human ES cell colonies

(Supplementary Fig. 2). This distinct profile of FGF receptor expression could not be attained from gene expression screens performed with heterogeneous bulk human ES cell cultures^{14,16}, and has therefore been overlooked. We suggest that FGF receptors may be important to human ES cell culture systems, but are not dominantly expressed by the self-renewing, pluripotent human stem cells.

Using immunocytochemistry, we observed that *FGFR1* expression was limited to cells surrounding the Oct4⁺ human ES cell colonies (Fig. 2a). In contrast, IGF1R was expressed exclusively by Oct4⁺ cells within human ES cell colonies (Fig. 2b). Furthermore, expression of *FGFR1* and IGF1R was found to be mutually exclusive within human ES cell cultures (Fig. 2c). This distinct expression pattern was observed in four independent human ES cell lines (H9, Fig. 2a–c, and H1, HES3 and CA1 in Fig. 2d, e). In addition, flow cytometric analysis indicated that human ES cell cultures expressing other markers associated with the pluripotent state, including SSEA4, Tra-1-81, Tra-1-60, but not SSEA1, also expressed IGF1R (Supplementary Fig. 3). IGF1R was also co-expressed with SSEA3 in several human ES cell lines (Supplementary Fig. 4). Unlike *FGFR1*, our results demonstrate that IGF1R expression correlates with pluripotent stem cell markers and hESC-CIC function, and thereby underscores both the uniqueness and general importance of the IGF-II/IGF1R axis in human ES cell lines.

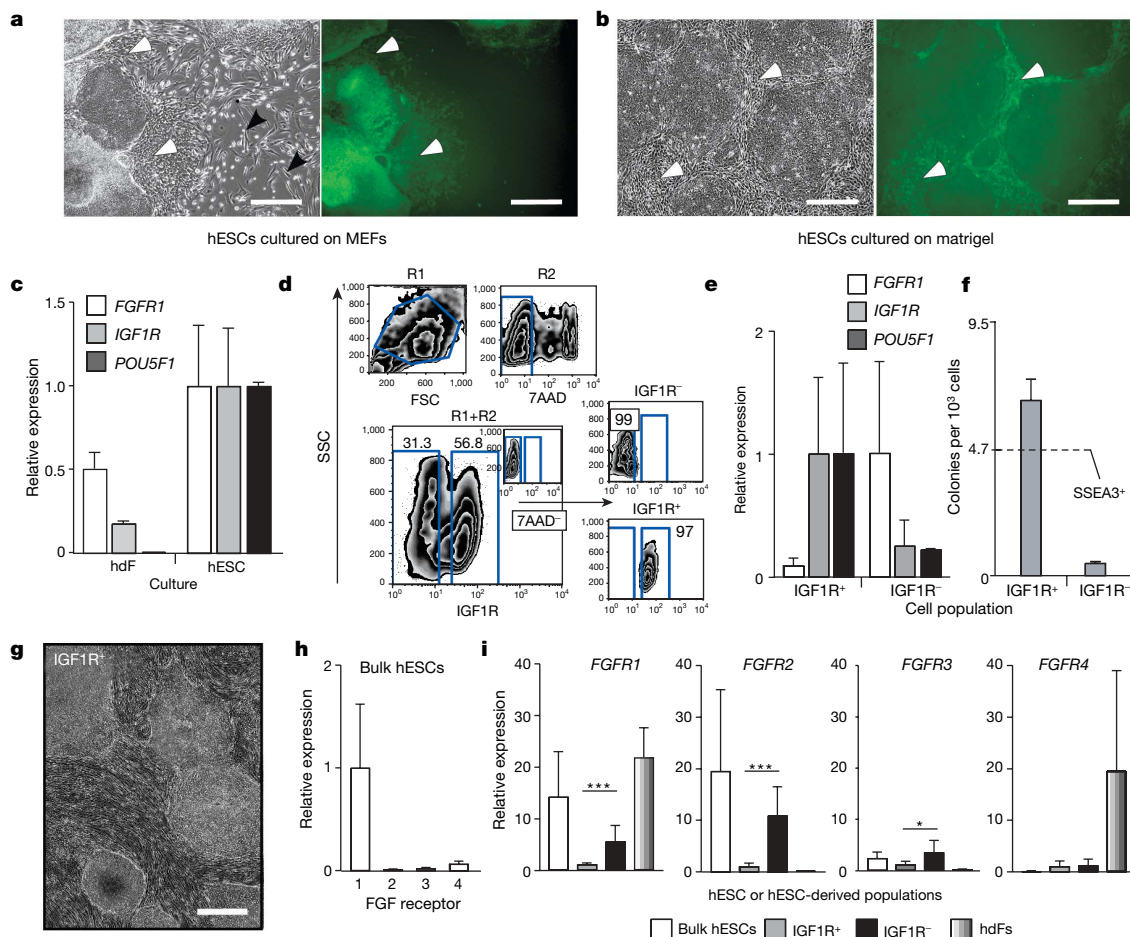


Figure 1 | Receptor expression reveals human ES cell heterogeneity.

a, b, Phase contrast and fluorescent images of GFP-labelled human ES cells (hESCs) forming hdFs (white arrowheads) cultured on MEFs (black arrowheads) (**a**) or matrigel. **c**, qPCR analysis of *FGFR1*, *IGF1R* and *POU5F1* (*Oct4*) expression in hdFs relative to human ES cell cultures ($n = 2$). **d**, Representative FACS plot of IGF1R⁺ 7-aminoactinomycin D (7AAD)⁻ human ES cell isolation. Gates (R1 and R2) and post-sort purity are shown. FSC, forward scatter; SSC, side scatter. **e**, qPCR analysis of *FGFR1*, *IGF1R*

and *POU5F1* expression relative to IGF1R⁺ in FACS isolated cells.

f, Clonogenic output of IGF1R⁺ or IGF1R⁻ FACS isolated cells after 6 days of culture. **g**, Day 6 colonies generated from IGF1R⁺ sorted cells. **h, i**, qPCR analysis of relative *FGFR1*–4 expression in bulk human ES cell cultures (**h**; $n = 3$) and IGF1R⁺ and IGF1R⁻ sorted human ES cell populations (**i**; $n = 3$) and hdFs (**i**; $n = 4$). All error bars indicate \pm s.e.m. Scale bars, 500 μ m. Asterisk, $P < 0.05$; triple asterisk, $P < 0.001$.

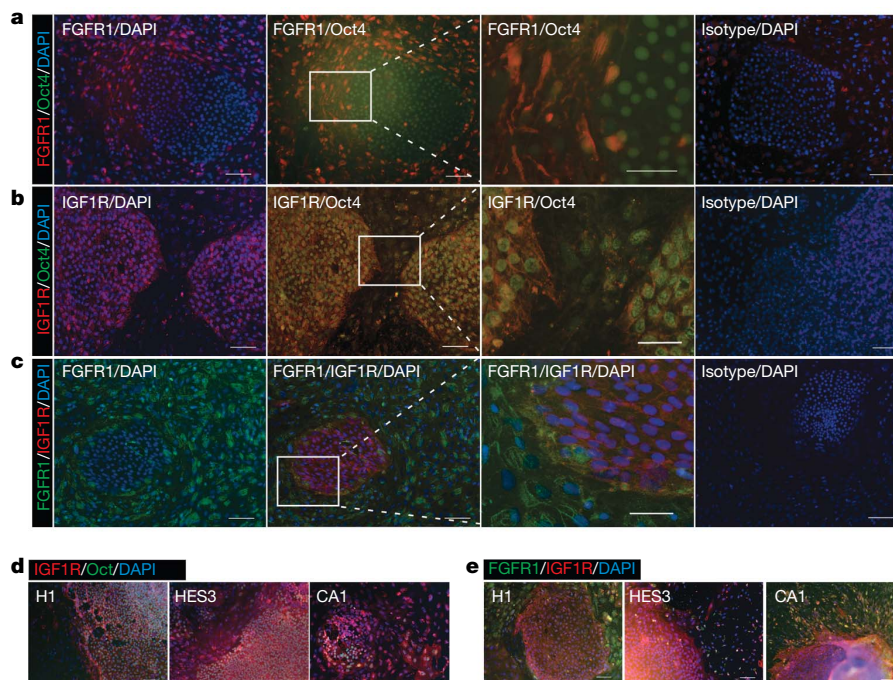


Figure 2 | Immunocytochemistry staining of undifferentiated human ES cell cultures. **a–c**, H9 cultures; **d, e**, H1, HES3 and CA1 cultures. Staining against FGFR1, Oct4 and IGF1R is shown and colours are indicated. Nuclei were counterstained with 4,6-diamidino-2-phenylindole (DAPI, blue).

Differential effects of IGF and FGF on human ES cell cultures

bFGF has an accepted, but undefined, role in the maintenance of human ES cell cultures^{12–15}. Because FGFR1 does not co-localize with Oct4 expression in human ES cell cultures, we sought to characterize functionally its role in contrast with IGF-II, the primary receptor of which, IGF1R, is expressed by Oct4⁺ hESC-CICs (Figs 1d–f and 2). We directly blocked both the IGF and FGF pathways through their respective receptors, IGF1R and FGFR1, in addition to antibody-mediated neutralization of the receptor ligands, IGF-II and bFGF. IGF1R blocking impeded human ES cell expansion as seen by a reduction in overall cell number. In contrast, inhibiting the FGF receptor had little effect on cell number, but increased differentiation (Fig. 3a–c). Long-term parallel analysis of both inhibitory conditions demonstrated that expression of SSEA3 was reduced to almost undetectable levels after 4 weeks (Fig. 3b, c). Blocking IGF1R caused a loss of SSEA3-expressing cells that corresponded to the loss of total cells (Fig. 3b), unlike FGF receptor inhibition, which had little to no effect on total cell number (Fig. 3c). Three weeks of continued neutralization of IGF-II and bFGF (insets of Fig. 3b, c respectively) yielded similar results compared to inhibiting their respective receptors. These effects were also observed using SSEA4 as a marker for human ES cell in two independent cell lines (Supplementary Figs 4 and 5). Patterns of FGFR1 expression combined with human ES cell differentiation in response to FGF inhibition suggest that FGF primarily has an indirect role in human ES cell maintenance. In contrast, correlation of IGF1R expression with the human ES cell phenotype collectively with impeded human ES cell expansion in response to IGF inhibition demonstrates IGF's direct role in human ES cell preservation, distinct from FGF.

We attempted to rescue blocked or inhibited cultures by compensating with high exogenous concentrations of the opposing ligand (Fig. 3d, e). bFGF had no positive effect on cumulative SSEA3⁺ cell number in the presence of the IGF1R blocking antibody (Fig. 3d). However, IGF-II completely rescued FGFR inhibition in terms of cumulative SSEA3⁺ cells (Fig. 3e). To delineate the effects of inhibiting the IGF and FGF pathways, we examined cell expansion per passage (expansion index), apoptosis in human ES cell and

a–c, The boxed area in the second column is magnified in the third column; isotype controls are shown in the final (right) column. Scale bars, 100 μm except for the third column of **a–c**, which are 50 μm. All human ES cells were cultured on matrigel with MEF-CM.

non-human ES cell populations, as well as differentiation. IGF1R inhibition significantly decreased human ES cell expansion per passage (Fig. 3f) resulting from a specific and significant increase in human ES cell (SSEA3⁺) apoptosis (Fig. 3g, h). This corresponds to a reduced mitotic index and increased apoptosis observed in other cell types known to be dependent on IGF1R signals^{17,18}. Accordingly, these data establish that human ES cells are also dependent for survival on signals through IGF1R. Although inhibition of FGFR1 had no significant influence on apoptosis (Fig. 3g, h), loss of the human ES cell phenotype corresponded to differentiation, as shown by increased SSEA3[−] cells (Fig. 3a, c) and by a significant increase in default neural differentiation compared to both control and blocking of IGF1R (Fig. 3i)^{19,20}. These results indicate that IGF-II specifically targets the colony-forming ES cell population in human ES cell cultures by directly affecting human ES cell expansion through increased proliferation and prevention of apoptosis. bFGF seems to act as a maintenance factor where, acting on the hF population, it indirectly prevents human ES cell differentiation yet has little influence on human ES cell proliferation.

IGF-II sustains self-renewal and pluripotency of ES cells

The addition of IGF-II alone in defined conditions was sufficient to maintain human ES cell developmental potential beyond 12 passages (Fig. 4a), as evidenced by the formation of teratomas containing all three embryonic germ layers (Fig. 4b). As human ES cell cultures are functionally heterogeneous with only a rare sub-fraction possessing clonogenic and culture regeneration capacity upon re-plating and serial passage⁶, factors acting on these hESC-CICs are the controllers of overall human ES cell self-renewal. To examine quantitatively and directly the functional effects of the IGF-II/IGF1R axis on clonogenic capacity and self-renewal, a novel clonogenic assay⁶ using a primary and secondary plating strategy for hESC-CICs was used (Fig. 4c). IGF1R interference reduced clonogenic capacity and self-renewal upon secondary re-plating compared to that of vehicle alone, whereas the addition of IGF-II enhanced both of these properties (Fig. 4d). The downstream effect of IGF1R antibody blocking was seen as a large decrease in Akt phosphorylation (Fig. 4e) and

downregulation of IGF1R on the cell surface (Supplementary Fig. 6), suggesting that the canonical IGF signalling pathway regulates functional properties of human ES cells. Taken together, these results demonstrate that IGF-II alone can sustain expansion of pluripotent human ES cells in long-term culture, and that the IGF signalling pathway via Akt activation is critical to human ES cell self-renewal.

hdFs produce supportive niche factors

IGF-II supplementation is not required in the absence of MEF influence to maintain human ES cells^{15,21}, suggestive of endogenous ligand production. Analysis of mouse and human IGF-II in MEF-CM and defined basal human ES cell media before and after human ES cell exposure revealed an increase in human IGF-II from undetectable to equivalent levels in both cases (Fig. 4f). For MEF-CM, mouse IGF-II was replaced completely by an approximately tenfold greater concentration of human IGF-II upon human ES cell exposure (Fig. 4f).

These data indicate that conventional human ES cell cultures can produce adequate amounts of endogenous IGF-II. As bFGF is important in maintaining human ES cells, we examined its influence on endogenous IGF-II production in both hdF and bulk human ES cell cultures. Inhibition of FGF, and addition of exogenous bFGF, induced opposite and significant effects on expansion of hdF cultures respectively (Fig. 4g). Moreover, IGF-II transcript in hdFs greatly increased in response to bFGF (Fig. 4h). Accordingly, when human ES cell cultures were exposed to bFGF for 24 h, a fourfold increase in fully processed 7.5-kDa IGF-II in serum-free media (Fig. 4i) was observed. These data suggest that the production of endogenous IGF-II, and possibly other human ES cell maintenance factors, is due in part to the influence of bFGF through induction of expression in, and maintenance of, FGF-responsive hdFs.

To identify other factors that could be produced as endogenous human ES cell regulators, we used a proteomic screen of defined

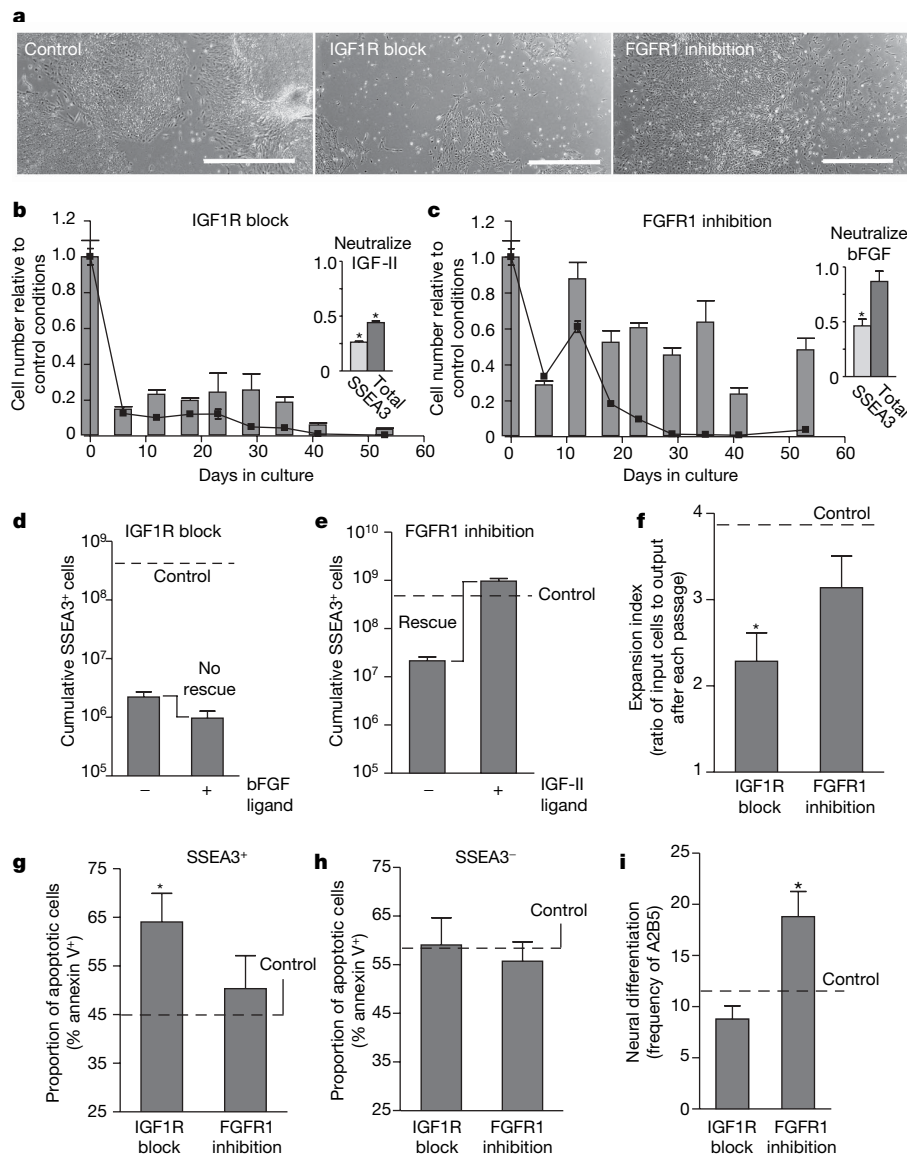


Figure 3 | Differential IGF and FGF effects on human ES cells. **a**, Phase images of human ES cell cultures treated with the indicated conditions. Scale bars, 500 μ m. **b**, **c**, Total (grey bars) and SSEA3⁺ (black squares) live human ES cells per well with the IGF1R blocking antibody (**b**) and day 21 IGF-II neutralization (inset), and the FGFR1 inhibitor SU5402 (**c**) and day 21 bFGF neutralization (inset). Both **b** and **c** were obtained in parallel and shown relative to control culture ($n = 4$). **d**, **e**, Rescue of IGF blockade with bFGF (**d**) or FGF blockade with IGF-II (**e**). Graphs indicate cumulative live

SSEA3⁺ human ES cells on day 53. **f**, Proliferative index of human ES cells with IGF or FGF blockade ($n = 8$). **g**, **h**, Apoptosis by annexin V staining on live (7AAD⁻) SSEA3⁺ (**g**) or SSEA3⁻ (**h**) human ES cells with IGF or FGF blockade, relative to control (MEF-CM) culture. Frequency is shown as mean \pm s.d. ($n = 3$). **i**, A2B5 expression in human ES cells with IGF or FGF blockade ($n = 4$). All graphs show mean \pm s.e.m. except where indicated. Asterisk, $P < 0.05$.

media conditioned by human ES cells. Aside from IGF, TGF- β was the most predominant growth factor family identified in hESC-CM (Supplementary Table 2). Consistent with FGF receptor inhibition, inhibition of TGF- β signalling increased differentiation (Fig. 5a) without affecting cell number (Fig. 5b), apoptosis, or cell cycle status in either the SSEA3⁺ (Fig. 5c) or SSEA3⁻ (Fig. 5d) human ES cell fractions. Treated human ES cell cultures rapidly lost the SSEA3⁺ population (Fig. 5e) and subsequent pluripotent potential *in vivo* (Fig. 5f, g) after 30 days of TGF- β inhibition. TGF- β 1 was highly expressed by hdFs but not by the IGF1R⁺ cells enriched for hESC-CICs (Fig. 5h). Moreover, TGF- β 1 expression was shown to be induced by FGF in human ES cell cultures²². Inhibition of TGF- β

caused a reduction of SMAD2/3 phosphorylation in human ES cells to levels similar to hdFs (Fig. 5i), corresponding to the observed morphological and phenotypic differentiation of human ES cells (Figs 5a, e). Consistent with their supportive role, these data suggest that hdFs produce TGF- β factors required by human ES cells to maintain an undifferentiated state. Moreover, *in vitro* and *in vivo* data signify that, unlike IGF but very similar to FGF, specific inhibition of TGF- β signals through SMAD2/3 results in loss of the human ES cell phenotype and pluripotency through induction of human ES cell differentiation, while not affecting survival or self-renewal of human ES cells. These results provide another example of FGF's role in the stem cell niche indirectly regulating pluripotent hESC-CICs.

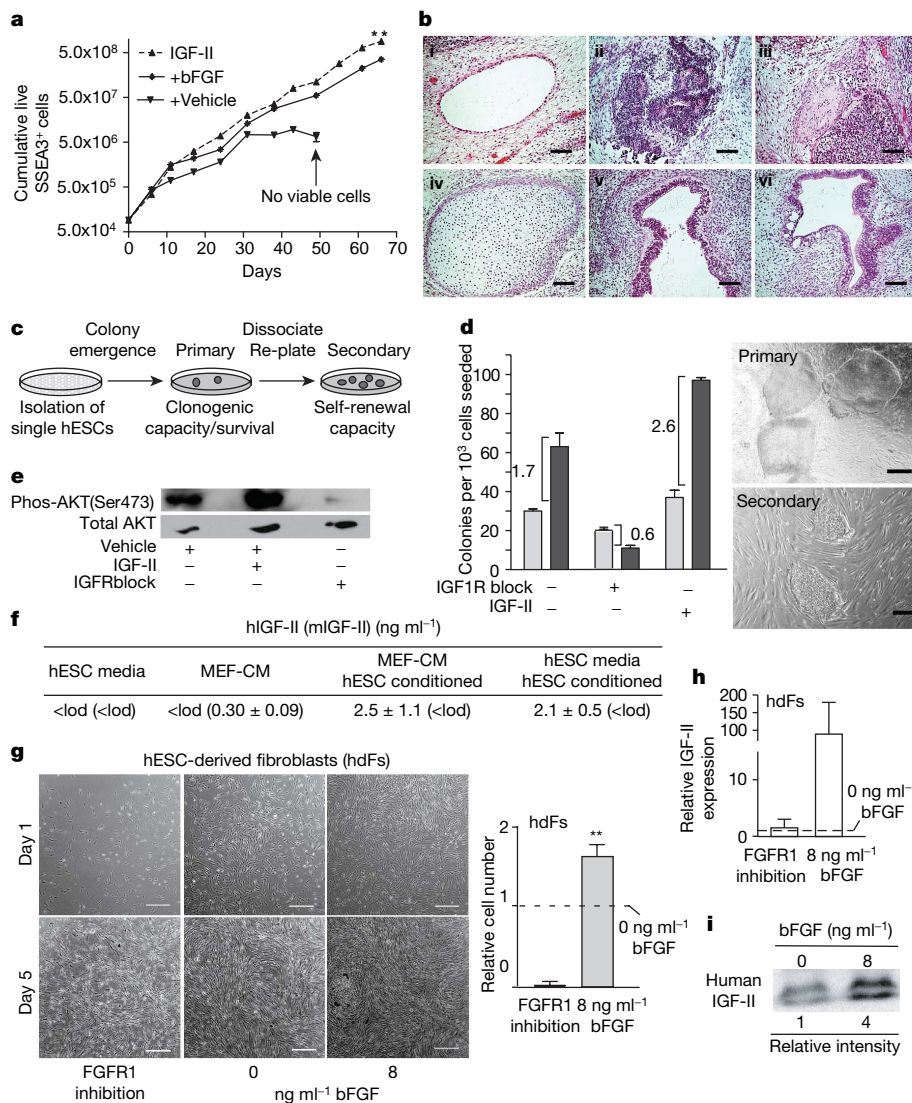


Figure 4 | Function of IGF and its production in human ES cell cultures.

a, Cumulative live SSEA3⁺ human ES cells in the presence of IGF-II alone or under control conditions (vehicle or bFGF alone) for 70 days. Double asterisk, developmental potential assessed. **b**, Sections from a 6-week teratoma formed from IGF-II cultured human ES cells (double asterisk in **a**, developmental potential assessed). Ectoderm: (i) cystic epithelium, (ii) neural rosettes; mesoderm: (iii) bone, (iv) cartilage; and endoderm: (v) gut-like glands with goblet cells, (vi) columnar epithelium. Scale bars, 50 μ m. **c**, Schematic of clonogenic and self-renewal capacity assay of sorted human ES cells (hESCs). **d**, Clonogenic capacity of SSEA3⁺ cells under indicated treatments at day 11 in primary (grey) plating and day 3 colony numbers in secondary (black) plating. Error bars indicate s.d. ($n = 3$). Human ES cell

colonies in primary and secondary plating of sorted human ES cells is shown (right). Scale bars, 500 μ m. **e**, Western blot showing Akt phosphorylation status in human ES cell cultures with IGF1R blocked for 48 h or stimulated with IGF-II for 30 min. **f**, Concentration of free human or mouse IGF-II in the indicated human ES cell media. Lod = limit of detection. **g**, Images (left) and relative cell numbers (right) of hdF cultures treated with FGFR1 inhibitor ($n = 2$), 8 ng ml⁻¹ bFGF ($n = 4$) or 0 ng ml⁻¹ bFGF ($n = 4$) (right, dashed line). Scale bars, 200 μ m. Double asterisk, $P < 0.01$. **h**, qPCR analysis of IGF-II expression in hdF cultures relative to 0 ng ml⁻¹ bFGF (dashed line) ($n = 2$). **i**, Western blot of IGF-II in hESC-CM. All graphs show mean \pm s.e.m.

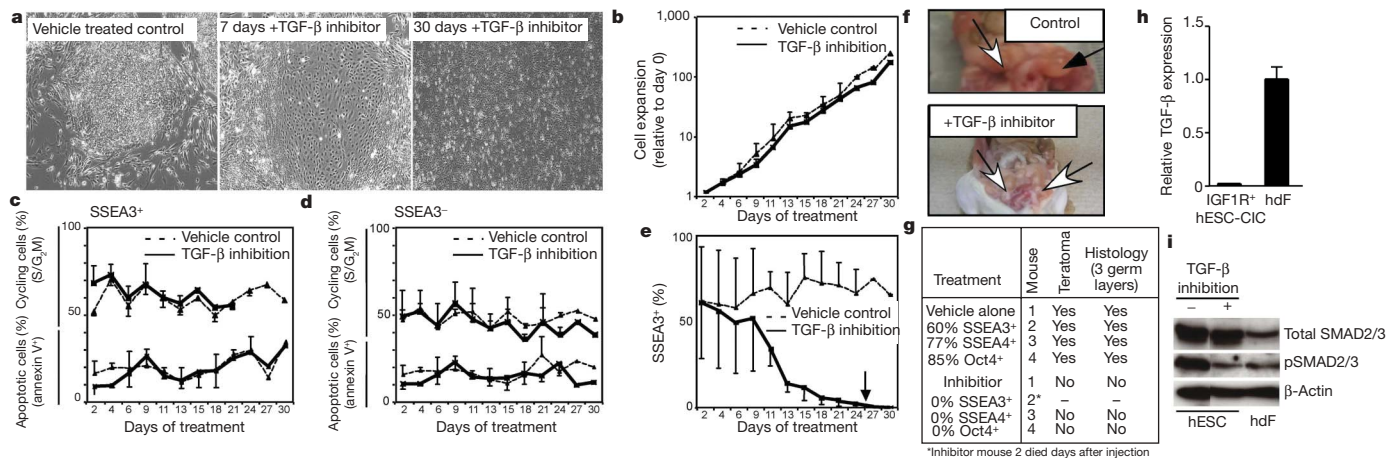


Figure 5 | FGF-induced TGF- β signals required for human ES cell pluripotency. **a**, Effect of TGF- β signal inhibition on human ES cell culture cell morphology. **b–e**, Human ES cell culture with and without TGF- β inhibition. Relative cell expansion (**b**), cell cycling and apoptotic status of SSEA3⁺ (**c**) and SSEA3[−] (**d**) cell subsets, and frequency of SSEA3⁺ cells (**e**) is shown. The arrow in **e** indicates complete loss. **f**, Teratoma images 8 weeks after injection of human ES cells maintained in the absence (top) or presence

(bottom) of SB431542. Teratoma, black arrow; normal testis, white arrow. **g**, Summary of teratoma formation from human ES cell cultures maintained for 30 days with or without TGF- β inhibition. **h**, TGF- β 1 expression in IGF1R-expressing FACS isolated human ES cells (hESCs; unsorted 45% IGF1R⁺) enriched for CICs compared to hdFs. **i**, Western blot of SMAD2/3 in control (-) and TGF- β inhibited (+) human ES cell cultures compared to hdFs; baseline. Error bars indicate s.e.m., $n = 4$ for each.

Discussion

Using proteomic identification, fractionation of human ES cell cultures and receptor localization combined with functional *in vitro* and *in vivo* assays to characterize human ES cell properties, we reveal a previously unappreciated but essential cellular interplay that establishes paracrine signalling as being required for self-renewal and pluripotency of human ES cells. On the basis of our results, we propose a model to describe the cellular interactions and signalling required to maintain human ES cells in culture (Fig. 6). Our study exposes a central and direct role for the IGF-II/IGF1R axis on human ES cells, while demonstrating that the action of FGF on human ES cells is primarily indirect. IGF has been implicated in clonal outgrowth of haematopoietic stem cells^{23,24}. In addition, IGF1R and IGF-II (but not IGF-I) transcripts are expressed in both human and mouse blastocysts²⁵, and IGF1R-mediated signalling is required for survival of the mouse inner cell mass²⁶. These previous studies, together with our current observations, suggest that the IGF-II/IGF1R axis has a universal role in understanding stem cell dependence on the niche. Our current findings clarify the role of FGF in maintaining human ES cell cultures, where FGF induces production

of supportive factors, including TGF- β factors and IGF-II, from the human ES-cell-derived fibroblast-like cells. Accordingly, FGF establishes the appropriate equilibrium of paracrine signalling among human ES cells and their supportive niche. Our results demonstrate that inhibition of FGF and TGF- β signals instructs human ES cell differentiation that is balanced by self-renewal signals controlled by IGF. Establishment of this paracrine signalling network requires acknowledgement of the previously unappreciated self-sustaining nature of human ES cell culture development. Obligatory differentiation of human ES cells to hDFs is central to establishing this signalling network among IGF, FGF and TGF- β , and thereby provides the framework in which human ES cells are instructed to self-renew or differentiate.

Although the concept that a regulatory niche creates an essential and supportive microenvironment for stem cells has been well established in several organisms and lineages¹, our study provides the first evidence that human ES cells possess the ability to autologously generate a stem cell niche (hdFs) *in vitro* despite the removal from their *in vivo* blastocytic microenvironment. Unlike adult stem cells, where the physical location is anatomically fixed *in vivo*, our findings outline an *in vitro* model that allows for the separation of the stem cells (IGF1R⁺, human ES cells) from the cells comprising the niche (FGFR1⁺, hdFs). The simplicity of this *in vitro* system provides an opportunity to elucidate the fundamental elements tightly controlling human stem cells within their niche that are not exclusive to pluripotent human ES cells, but instead more broadly applicable to multiple stem cell types.

METHODS SUMMARY

Human ES cell lines H1, H9, HES3 and CA1 were maintained in feeder-free culture in MEF-CM³ or, in the case of H1 and H9, in the absence of conditioning with 36 ng ml⁻¹ bFGF as previously described²¹. hdFs were established and cultured as previously described⁶. Expression of pluripotent human ES cell markers, specifically SSEA3, SSEA4, Tra-1-60, Tra-1-81 and Oct4 (POU5F1), in addition to FGFR1-4 and IGF1R expression in bulk human ES cell cultures, FACS isolated human ES cell populations and hdF cultures, were characterized by flow cytometry, immunocytochemistry and/or quantitative reverse transcriptase PCR (qPCR). FACS isolation and quantitative clonogenic analysis of human ES cell populations based on IGF1R expression was performed as previously described⁶. Interference with FGF, IGF and TGF- β signalling in human ES cell cultures was accomplished by ligand neutralization or receptor blocking through addition of antibodies (clones 1H7 and α -IR3 for IGF1R) or chemical inhibitors (SU5402 for FGFR and SB431542 for SMAD2/3), where their effects were assessed in parallel with appropriate controls (immunoglobulin isotype match for antibodies and

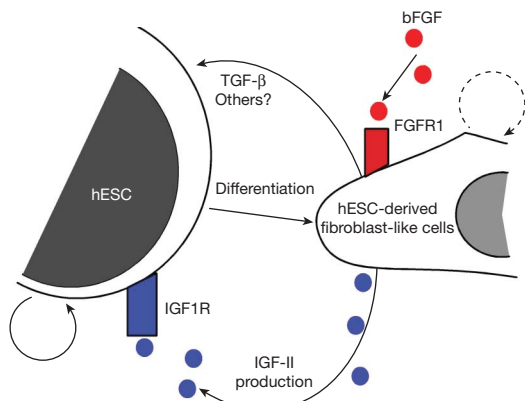


Figure 6 | Proposed model of human ES cell paracrine regulation. To maintain culture homeostasis within its autologously produced niche, human ES cells (hESCs) spontaneously and continuously differentiate into hdFs, providing a continuous source of endogenous human ES cell supportive factors, including IGF-II and a host of TGF- β family of factors and other ligands.

DMSO for chemical inhibitors). Effects of interference on human ES cell cultures were monitored by analysing pluripotent human ES cell marker expression, total viable cell number, apoptosis (annexin V staining), cell cycle (propidium iodide staining) and default differentiation (A2B5 expression). Human ES cells were cultured in defined conditions using IGF-II (30 ng ml⁻¹) alone for greater than 70 days at which point *in vivo* developmental capacity was assessed by the teratoma assay through intratesticular injection into NOD-SCID mice²¹. Unbound IGF-II was quantified using mouse- or human-specific enzyme-linked immunosorbent assay (ELISA) whereas total IGF-II was determined by western blotting. Akt and SMAD2/3 phosphorylation status were analysed by western blotting in human ES cells after 48 h of indicated interference treatment. IGF-II secretion/expression, and TGF- β 1 expression in hEF and/or human ES cell cultures in the absence or presence of FGF, were analysed by western blot and/or qPCR. Further details of all experiments are provided in the Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 April; accepted 18 June 2007.

Published online 11 July 2007.

- Scadden, D. T. The stem-cell niche as an entity of action. *Nature* **441**, 1075–1079 (2006).
- Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
- Xu, C. *et al.* Feeder-free growth of undifferentiated human embryonic stem cells. *Nature Biotechnol.* **19**, 971–974 (2001).
- Ludwig, T. E. *et al.* Derivation of human embryonic stem cells in defined conditions. *Nature Biotechnol.* **24**, 185–187 (2006).
- Hoffman, L. M. & Carpenter, M. K. Characterization and culture of human embryonic stem cells. *Nature Biotechnol.* **23**, 699–708 (2005).
- Stewart, M. H. *et al.* Clonal isolation of hESCs reveals heterogeneity within the pluripotent stem cell compartment. *Nature Methods* **3**, 807–815 (2006).
- Ying, Q. L., Nichols, J., Chambers, I. & Smith, A. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115**, 281–292 (2003).
- DeChiara, T. M., Efstratiadis, A. & Robertson, E. J. A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting. *Nature* **345**, 78–80 (1990).
- Kauma, S. W. Cytokines in implantation. *J. Reprod. Fertil., Suppl.* **55**, 31–42 (2000).
- Heyner, S. Growth factors in preimplantation development: role of insulin and insulin-like growth factors. *Early Pregnancy* **3**, 153–163 (1997).
- Sperger, J. M. *et al.* Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. *Proc. Natl Acad. Sci. USA* **100**, 13350–13355 (2003).
- Levenstein, M. E. *et al.* Basic fibroblast growth factor support of human embryonic stem cell self-renewal. *Stem Cells* **24**, 568–574 (2006).
- Xu, R. H. *et al.* Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. *Nature Methods* **2**, 185–190 (2005).
- Dvorak, P. *et al.* Expression and potential role of fibroblast growth factor 2 and its receptors in human embryonic stem cells. *Stem Cells* **23**, 1200–1211 (2005).
- Xu, C. *et al.* Basic fibroblast growth factor supports undifferentiated human embryonic stem cell growth without conditioned medium. *Stem Cells* **23**, 315–323 (2005).
- Rosler, E. S. *et al.* Long-term culture of human embryonic stem cells in feeder-free conditions. *Dev. Dyn.* **229**, 259–274 (2004).
- Baserga, R., Peruzzi, F. & Reiss, K. The IGF-1 receptor in cancer biology. *Int. J. Cancer* **107**, 873–877 (2003).
- Hofmann, F. & Garcia-Echeverria, C. Blocking the insulin-like growth factor-I receptor as a strategy for targeting cancer. *Drug Discov. Today* **10**, 1041–1047 (2005).
- Tropepe, V. *et al.* Direct neural fate specification from embryonic stem cells: a primitive mammalian neural stem cell stage acquired through a default mechanism. *Neuron* **30**, 65–78 (2001).
- Vallier, L., Reynolds, D. & Pedersen, R. A. Nodal inhibits differentiation of human embryonic stem cells along the neuroectodermal default pathway. *Dev. Biol.* **275**, 403–421 (2004).
- Wang, L., Li, L., Menendez, P., Cerdan, C. & Bhatia, M. Human embryonic stem cells maintained in the absence of mouse embryonic fibroblasts or conditioned media are capable of hematopoietic development. *Blood* **105**, 4598–4603 (2005).
- Greber, B., Lehrach, H. & Adjaye, J. Fibroblast growth factor 2 transforming growth factor β signaling in mouse embryonic fibroblasts and human ESCs (hESCs) to support hESC self-renewal. *Stem Cells* **25**, 455–464 (2007).
- Huang, S. & Terstappen, L. W. Formation of haematopoietic microenvironment and haematopoietic stem cells from single human bone marrow stem cells. *Nature* **360**, 745–749 (1992).
- Huang, S. & Terstappen, L. W. Formation of haematopoietic microenvironment and haematopoietic stem cells from single human bone marrow stem cells. *Nature* **368**, 664 (1994).
- Lighten, A. D., Hardy, K., Winston, R. M. & Moore, G. E. Expression of mRNA for the insulin-like growth factors and their receptors in human preimplantation embryos. *Mol. Reprod. Dev.* **47**, 134–139 (1997).
- Chi, M. M., Schlein, A. L. & Moley, K. H. High insulin-like growth factor 1 (IGF-1) and insulin concentrations trigger apoptosis in the mouse blastocyst via down-regulation of the IGF-1 receptor. *Endocrinology* **141**, 4784–4792 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements S.C.B. is supported by a CIHR Canada Graduate Scholarship doctoral award; M.H.S. by a postgraduate scholarship award from the Stem Cell Network and CIHR Canada Graduate Scholarship doctoral award; and M.Bh. by the Canadian Chair Program who holds the Canada Research Chair in human stem cell biology and Michael G. DeGroote Chair in Stem Cell Biology. This work was supported by a grant from the Ontario Research and Development Challenge Fund (ORDCF) to G.L. and by CIHR and NCIC to M.Bh. We also are grateful for the help of L. Gallacher and R. Mondeh with culture assistance, the Krembil Centre at the Robarts and M. Sibly and J. Trowbridge for useful suggestions, and A. Nagy, J. Rossant, M. Gertsenstein, K. Vinterstein, M. Mileikovsky and J. Draper for providing the CA1 human ES cell line.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.B. (mbhatia@mcmaster.ca).

METHODS

General human ES cell culture. Human ES cell lines H1, H9, HES3 and CA1 were maintained in feeder-free culture in MEF-CM³ or, in the case of H1 and H9, in the absence of conditioning with 36 ng ml⁻¹ bFGF²¹, SR-FGF culture. Human ES cells were plated on matrigel-coated (BD Biosciences) plates, the media was changed daily, and the cells were passaged every 5–7 days through dissociation with 200 U ml⁻¹ collagenase IV (Invitrogen). MEF-CM with corresponding treatments was changed daily and human ES cells were 1:2 passaged weekly by dissociation with 200 U ml⁻¹ of collagenase IV (Invitrogen). Recombinant human IGF-II (Upstate) was reconstituted according to the manufacturer's instructions and stored at 10 µg ml⁻¹ in PBS with 3% BSA at -30 °C. IGF-II was added fresh daily and used at 30 ng ml⁻¹ concentration. All experiments were performed on H1 and H9 human ES cells. Immunocytochemistry and flow cytometry were performed on H1, H9, HES3 and CA1.

hdF culture. hdF cultures were established as previously described⁶. Briefly, human ES cell cultures with a high proportion of fibroblast-like cells were treated for 20 min with collagenase IV. Collagenase IV was removed and wells were rinsed with KO-DMEM (Invitrogen) to collect the fibroblast-like cells. Supernatant containing the cells was centrifuged at 450g for 5 min and the cell pellet was re-suspended in MEF-CM plus 8 ng ml⁻¹ bFGF or SR-bFGF and transferred to fresh matrigel-coated wells. The media was changed every other day and hdFs were passaged when confluent in the same manner as for human ES cells.

Reverse transcription and qPCR. Total RNA was extracted from human ES cells, human ES-cell-derived fibroblast-like cells (hdFs) and sorted cells using the Qiagen AllPrep DNA/RNA mini kit (Qiagen number 80204). cDNA was generated with 1–5 µg of total RNA using the first-strand cDNA synthesis kit (Amersham Biosciences). Quantitative PCR was performed using Platinum SYBR Green qPCR Super Mix-UDG on an Mx3000P qPCR System (Stratagene) or a Chromo4 Real-Time PCR detector (BioRad) according to the manufacturer's instructions (Invitrogen). Quantitative PCR amplifications were performed using the following conditions: 95 °C, 10 min and 40 cycles 95 °C, 30 s; 60 °C, 1 min; 72 °C, 30 s. Analysis and comparison of FGFR1, FGFR2, FGFR3 and FGFR4 was performed using primers and conditions previously published¹⁴. All data were normalized to GAPDH.

Gene sequence amplified product. IGF1R-F, 5'-CCAAAGACAAAATC-CCCATC-3' 99 bp; IGF1R-R, 5'-CTTTCTCCCCACCACACAC-3'; FGFR1-F, 5'-GGCTACAAGGTCCGTATGTC-3' 119 bp; FGFR1-R, 5'-TGGTATGTGTGTTGATGCTG-3'; POU5F1-F, 5'-CTGAAGCAGAAGAGGATCAC-3' 366 bp; POU5F1-R, 5'-GACCACATCCTTCTCGAGCC-3'; IGF-II-F, 5'-TTGCTCTACCCACCAAGAC-3' 150 bp; IGF-II-R, 5'-GATGGAACCTGATGGAAACG-3'; GAPDH-F, 5'-TGTCATCAATGGAAATCCCATCACC-3' 325 bp; GAPDH-R, 5'-CATGAGTCTCCACGATACCAAA-3'; TGF-β1-F, 5'-GCGTGTCTAATGGTGGAAC-3' 221 bp; TGF-β1-R, 5'-GCTGAGGTATCGCAGGAAT-3'.

Characterization of human ES cells by flow cytometry. Single-cell suspensions were made from undifferentiated human ES cells and general staining procedures were previously described²¹. Stained cells were analysed on a FACSCalibur using Cell Quest software (BD Biosciences). Unconjugated antibody against SSEA3 (MC-631, Develop Studies Hybridoma Bank), SSEA4 (MC-813-70, Develop Studies Hybridoma Bank), SSEA1 (MC480, Develop Studies Hybridoma Bank), Tra-1-60 (Chemicon), Tra-1-81 (Chemicon) and Oct3 (BD Bioscience catalogue number 611203) or Oct3/4 (N-19 Santa Cruz SC-8628) were used to viably identify human ES cells. The IGF1R surface expression was assessed using the anti-IGF1R clone 1H7 conjugated to phycoerythrin (BD Pharmingen) according to the manufacturer's instructions, or 10 µg ml⁻¹ chicken anti-IGF1R (Abcam catalogue number ab32823) with Alexa Fluor 488 goat anti-chicken IgG (Molecular Probes). Finally, cells were also stained for viability based on exclusion of the 7-aminoactinomycin D (7-AAD) dye.

Human ES cell sorting, clonogenic and self-renewal assays. Human ES cells were dissociated and stained for the indicated marker as described above. Live human ES cells were identified by 7-AAD exclusion and selected for based on expression of either SSEA3 or IGF1R as indicated using a FACSVantage (BD Pharmingen) or FACS Aria (BD Pharmingen). Sorted cells were concentrated by centrifugation before being seeded into wells containing matrigel plus irradiated hdFs⁶. For clonogenic capacity, colony numbers were monitored every 2 to 3 days after sorting. After 11 days, colonies were dissociated, low dilutions were replated in the same conditions, and colonies were counted 3 days later to assess self-renewal capacity of individual human ES cell clones. For qPCR analysis, sorted cells were pelleted and RNA was extracted as described above.

Immunocytochemical detection of IGF1R, FGFR1 and Oct4 expression in human ES cell and hdF cultures. Human ES cells and hdFs were cultured as described above until 70–80% confluent in matrigel-coated 4-well chamber

slides (Lab Tek Permanox, VWR, catalogue number CA62407-320). Cells were rinsed twice with PBS before fixation with 4% paraformaldehyde, followed by receptor staining before permeabilization with 0.2% Triton X-100 or 100% methanol and staining for Oct4. Cells were blocked with 10% normal rabbit serum plus 10% normal goat serum plus 5% BSA at room temperature. Antibodies used were 10 µg ml⁻¹ chicken anti-IGF1R (Abcam ab32823), 2 µg ml⁻¹ mouse anti-FGFR1 clone VBS1 (Chemicon MAB125), 1 µg ml⁻¹ each of anti-FGFR2, -FGFR3 and -FGFR4 rabbit polyclonal antibodies (Santa-Cruz Biotechnology SC-122, SC-123, S-124 respectively), 3 µg ml⁻¹ goat anti-Oct3/4 (N-19) (Santa Cruz Biotechnology, catalogue number SC-8628) and 5 µg ml⁻¹ mouse anti-Oct3 (BD Bioscience catalogue number 611203). Cells were incubated with primary antibodies followed by secondary detection with Alexa Fluor 488 or 594 goat anti-mouse IgM (Molecular Probes), Alexa Fluor 488 or 594 rabbit anti-goat IgG (Molecular Probes), and/or Alexa Fluor 488 or 594 goat anti-chicken IgG (Molecular Probes) at 3 µg ml⁻¹ or Alexa Fluor 488 rabbit anti-mouse IgG (Molecular Probes) at 5 µg ml⁻¹. Chamber slides were mounted and counterstained using VECTASHIELD HardSet Mounting Medium with DAPI (Vector Labs). Slides were examined using an Olympus IX81 or IX51 microscope and images were captured with a Photometrix Cool Snap HQ² camera using *in vivo* version 3.1.2 (Photometrix) software or Image-Pro 3DA version 6.0. Images were pseudo-coloured and analysed using Image-Pro 3DA version 6.0.

Inhibition of IGF, FGF and TGF-β in cell cultures. The effect of inhibiting IGF signalling via IGF1R has been studied extensively using well characterized blocking antibodies¹⁸. In experiments blocking IGF1R signals we used either anti-IGF1R antibody clone 1H7 (BD Pharmingen or Santa Cruz Biotechnology) or clone alpha-IR3 (Oncogene) at a final concentration of 2 µg ml⁻¹ in human ES cell culture. A concentration of 4 µg ml⁻¹ of antibody was also tested and displayed no increase in effect (data not shown). These represent 2–4 times the recommended effective concentration by the manufacturer. As a control, affinity-purified mouse IgG1-κ (BD Pharmingen) was used at the equivalent concentration. Antibody was added fresh to human ES cell media whenever it was changed. Inhibition of bFGF signals in human ES cells via FGFR1 has been previously demonstrated with the FGFR1-specific chemical inhibitor SU5042 (refs 14, 27). FGFR1 was inhibited with 20 µM SU5042 (Calbiochem) with an equivalent volume of DMSO used as the vehicle control. TGF-β signalling was inhibited using 10 µM of SB431542 (Sigma), a potent and selective chemical inhibitor of ALK4, ALK5 and ALK7 receptors inducing shutdown of TGF-β/activin signalling via SMAD2/3^{28,29}, or equivalent volume of DMSO (vehicle). bFGF and IGF-II neutralization was performed using 0.8 µg ml⁻¹ anti-recombinant human (rh)bFGF (Abcam ab12475) or 3 µg ml⁻¹ anti-rhIGF-II (R&D Systems number AF-292-NA) goat polyclonal antibodies respectively with an equivalent concentration of affinity pure goat IgG (R&D) as a control.

Expansion and apoptosis analysis. Cumulative cell counts and cell expansion was assessed as previously described²¹. Expansion index was calculated for multiple passages in the indicated condition as follows: (live cells collected per well at passage N)/(live cells plated per well from passage N–1). Apoptosis was assayed via annexin V staining using an apoptosis detection kit (BD Bioscience) according to manufacturer's recommendations. Forty-eight hours before analysis cells were treated with inhibitory condition or control, unless otherwise indicated. The levels of apoptosis within SSEA3⁺ and SSEA3⁻ cells were assessed by co-staining with anti-SSEA3 as mentioned previously. Briefly, human ES cell cultures dissociated and stained for SSEA3 (as indicated above) were washed twice and stained with annexin-V-phycoerythrin and 7-AAD following instructions from the manufacturer. Apoptotic cells were identified by gating the Annexin-V⁺ 7AAD⁻ fraction within SSEA3⁺ and SSEA3⁻ cell subpopulations.

Cell cycle analysis. To analyse the effect of the different culture conditions on the cycling status of SSEA3⁺ and SSEA3⁻ cells, DNA content was measured by means of propidium iodide (Sigma) staining. In brief, 0.5–1 × 10⁶ single cells were stained for SSEA3, washed in PBS containing 3% FBS and fixed in cold 70% ethanol for 2–3 h at 4 °C. After washing in 100 mM NaCl, cells were permeabilized and DNA stained with a solution containing 0.2 mg ml⁻¹ DNase-free RNase A (Sigma) and 0.02 mg ml⁻¹ propidium iodide prepared in 0.1% Triton X-100 in 100 mM NaCl. Cell cycle status was analysed in a FACSCalibur with CellQuest software and a minimum of 3 × 10⁴ cells were acquired at low flow rate (15–75 cells s⁻¹). On the basis of propidium iodide staining, G0/G1, S and G2/M phases of the cell cycle were identified. The G0/G1 peak consistently displayed a CV < 7%^{30,31}.

Teratoma formation and histological analysis. McMaster University's council on Animal Care approved of all animal protocols here. The formation and histological analysis of human ES-cell-derived teratomas was carried out as previously described²¹.

Karyotype. The karyotype analysis was done at the London Health Science Centre, Cytogenetics Laboratory (London, Ontario, Canada).

ELISA and western blot assays. Free mouse IGF-II was assessed using the Duo Set ELISA kit (R&D Systems). Free human IGF-II was determined with the Non-Extraction IGF-II ELISA kit (Diagnostic Systems Laboratories). Note that there was less than 5% cross-reactivity between the mouse- and human-specific IGF-II assays. Assays were performed according to the manufacturer's recommended protocol. Total IGF-II was assessed by western blot analysis performed on concentrated serum-free media either exposed to MEFs or human ES cells and using antibody clone S1F2 (Upstate) according to recommendations of the manufacturer.

For the effect of IGF blocking on Akt (Ser 473) phosphorylation, day 4 human ES cells were exposed to the IGF1R blocking antibody (1H7) or control mouse IgG for 48 h. Before harvesting, 30 ng ml⁻¹ of IGF-II was added to one of the two IgG control conditions for a 30 min activation. Approximately 40 µg of lysate from each condition was analysed by western blot using rabbit anti-phospho-Akt (Ser 473) (Cell Signaling Technologies catalogue number 9271), or for total Akt with rabbit anti-Akt (Cell Signaling catalogue number 9272). Primary antibodies were used according to the manufacturers' instructions. In both cases they were detected with goat anti-rabbit-HRP conjugate (Pierce) at a 1:7,500 dilution.

For assessment of phosphorylated SMAD2/3 by western blotting, cell lysates were prepared from equivalent cell numbers maintained short-term (~48 h) in conditions indicated under standard human ES cell or hF culture conditions. Approximately 30 µg of cell lysate was assessed using rabbit anti-phospho-SMAD2/3 (Cell Signaling technologies, 1:10,000) or total SMAD2/3 (Cell Signaling, 1:10,000). Membranes were washed and stained with HRP-conjugated goat anti-rabbit (BioRad, 1:30,000).

Cells were washed and lysed using the nuclear extraction kit with protease and phosphatase inhibitors (Active Motif) using the whole-cell lysis procedure. Samples were run on 10% or 12% SDS-PAGE and transferred to PVDF using a semi-dry transfer apparatus at 15 V for 1 h. Membranes were blocked with either 5% skim milk or 5% BSA in TBST and incubated with primary antibodies overnight at 4 °C. All secondary antibodies were conjugated to HRP and consequently detected using the enhanced chemiluminescence system (ECL) (Pierce). ECL signal was detected using X-ray film or by using an imaging detection and analysis station (Alpha-Innotech Corp). For western blot loading/total protein controls, membranes were either stripped using 250 mM glycine, pH 2.5, and re-probed as described, or an equivalent amount of sample was run and detected in parallel as previously described.

Statistical analysis. Results are expressed as mean ± s.d. or s.e.m. as indicated in the figure legends. Statistical significance was determined using two-way analysis of variance analysis or paired Student's *t*-test. Differences were reported as follows: asterisk, *P* < 0.05; double asterisk, *P* < 0.01; triple asterisk, *P* < 0.001.

27. Mohammadi, M. *et al.* Structures of the tyrosine kinase domain of fibroblast growth factor receptor in complex with inhibitors. *Science* **276**, 955–960 (1997).
28. Inman, G. J. *et al.* SB-431542 is a potent and specific inhibitor of transforming growth factor-β superfamily type I activin receptor-like kinase (ALK) receptors ALK4, ALK5, and ALK7. *Mol. Pharmacol.* **62**, 65–74 (2002).
29. Laping, N. J. *et al.* Inhibition of transforming growth factor (TGF)-β1-induced extracellular matrix with a novel inhibitor of the TGF-β type I receptor kinase activity: SB-431542. *Mol. Pharmacol.* **62**, 58–64 (2002).
30. Orfao, A. *et al.* Flow cytometry in the diagnosis of cancer. *Scand. J. Clin. Lab. Invest., Suppl.* **221**, 145–152 (1995).
31. Orfao, A. *et al.* Flow cytometry: its applications in hematology. *Haematologica* **80**, 69–81 (1995).

LETTERS

Rapid planetesimal formation in turbulent circumstellar disks

Anders Johansen¹, Jeffrey S. Oishi^{2,3}, Mordecai-Mark Mac Low^{1,2}, Hubert Klahr¹, Thomas Henning¹ & Andrew Youdin⁴

During the initial stages of planet formation in circumstellar gas disks, dust grains collide and build up larger and larger bodies¹. How this process continues from metre-sized boulders to kilometre-scale planetesimals is a major unsolved problem²: boulders are expected to stick together poorly³, and to spiral into the protostar in a few hundred orbits owing to a 'headwind' from the slower rotating gas⁴. Gravitational collapse of the solid component has been suggested to overcome this barrier^{1,5,6}. But even low levels of turbulence will inhibit sedimentation of solids to a sufficiently dense midplane layer^{2,7}, and turbulence must be present to explain observed gas accretion in protostellar disks⁸. Here we report that boulders can undergo efficient gravitational collapse in locally overdense regions in the midplane of the disk. The boulders concentrate initially in transient high pressure regions in the turbulent gas⁹, and these concentrations are augmented a further order of magnitude by a streaming instability^{10–12} driven by the relative flow of gas and solids. We find that gravitationally bound clusters form with masses comparable to dwarf planets and containing a distribution of boulder sizes. Gravitational collapse happens much faster than radial drift, offering a possible path to planetesimal formation in accreting circumstellar disks.

Planet formation models typically treat turbulence as a diffusive process that opposes the gravitational sedimentation of solids to a high density midplane layer in circumstellar disks^{7,13}. Recent models of solids moving in turbulent gas reveal that the turbulent motions not only mix them, but also concentrate metre-sized boulders in the transient gas overdensities⁹ formed in magnetorotational turbulence¹⁴, in giant gaseous vortices^{15,16}, and in spiral arms of self-gravitating disks¹⁷. Short-lived eddies at the dissipation scale of forced turbulence concentrate smaller millimetre-sized solids¹⁸.

Some simulations mentioned above^{9,11,12} were performed with the Pencil Code, which solves the magnetohydrodynamic equations on a three-dimensional grid for a gas that interacts through drag forces with boulders. Boulders are represented as superparticles with independent positions and velocities, each having the mass of a huge number of boulders but the aerodynamic behaviour of a single boulder. We have now further developed the Pencil Code to include a fully parallel solver for the gravitational potential of the particles (see Supplementary Information). The particle density is mapped on the grid using the 'triangular shaped cloud' assignment scheme¹⁹ and the gravitational potential of the solids is found using a fast Fourier transform method²⁰. This allows us, for the first time to our knowledge, to simulate the dynamics of self-gravitating solid particles in magnetized, three-dimensional turbulence.

We model a co-rotating, local box with linearized keplerian shear that straddles the protoplanetary disk midplane and orbits the young

star at a fixed distance. Periodic boundary conditions are applied. An isothermal equation of state is used for the gas, whereas the induction equation is solved under the ideal magnetohydrodynamic assumption of high conductivity. Magnetorotational instability¹⁴ drives turbulence in keplerian disks with sufficient ionization²¹, producing in our unstratified models turbulence with Mach number $Ma \approx 0.05$ and viscosity $\alpha \approx 10^{-3}$, a realistic value to explain observed accretion rates⁸. The ionization fraction in the dense midplanes of protoplanetary disks may be insufficient for the gas to couple with the magnetic field to drive magnetorotational instability²¹. In the Supplementary Information we therefore describe unmagnetized models as well.

Solid objects orbit the protostar with keplerian velocity v_K in the absence of gas drag. A radial pressure gradient partly supports the gas, however, so it orbits at sub-keplerian velocity v_g , with $\Delta v \equiv v_g - v_K < 0$. As a result, large (approximately metre-sized) solid objects are exposed to a strong headwind that causes them to drift radially inwards⁴ with a maximum drift velocity Δv . They also 'feel' gas drag as they fall towards the disk midplane in the effective gravity field of the star. A sedimentary midplane layer forms, with a width determined by a balance between settling and turbulent diffusion^{7,13}.

We present three types of models: (1) without self-gravity, with 128^3 zones and 2×10^6 particles, run for 100 orbits, to study the interplay between the streaming instability¹⁰ and concentration by transient high pressures; (2) with self-gravity and boulder collisions, with 256^3 zones and 8×10^6 particles run for 27 orbits, to study gravitational collapse; and (3) models with self-gravity but no magnetorotational turbulence (presented in the Supplementary Information). Magnetorotational turbulence is given 10 orbits to reach steady state before we turn on drag force and vertical gravity, to avoid any influence of the initial conditions on the sedimented midplane layer. We fix the global solids-to-gas bulk density ratio at the canonical galactic value of $\epsilon_0 = 0.01$, but two values of the radial drift are considered: low drift, with $\Delta v = -0.02c_s$, where c_s is the isothermal sound speed, and moderate drift, with $\Delta v = -0.05c_s$, depending on the assumed radial pressure support (values up to $\Delta v = -(0.2-0.5)c_s$ are possible¹³, but are not considered here).

For the simulations without self-gravity, we consider a fixed particle size parameterized by the dimensionless friction time $\Omega_K \tau_f = 1.0$, where Ω_K is the local keplerian rotation frequency and τ_f is the timescale over which gas and solids reach equal velocity. At an orbital distance $r = 5 \text{ AU}$, this corresponds to boulders of approximately one metre in diameter. Figure 1 shows the space-time topography of the sedimented midplane layer. The streaming instability increases the density of boulders in regions where they have already

¹Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany. ²Department of Astrophysics, American Museum of Natural History, 79th Street at Central Park West, New York, New York 10024-5192, USA. ³Department of Astronomy, University of Virginia, Charlottesville, Virginia 22904, USA. ⁴Canadian Institute for Theoretical Astrophysics, University of Toronto, 60 St George Street, Toronto, Ontario M5S 3H8, Canada.

been concentrated by transient high pressures⁹. Increasing radial pressure support, by changing $\Delta\nu = -0.02c_s$ to $-0.05c_s$, reduces the concentration (see Fig. 1 legend for explanation), although the local solids-to-gas density ratio still reaches 200.

Gravitational collapse of discrete solid objects produces virialized clusters unable to contract further²² in the absence of mechanisms to dynamically cool the cluster—that is, to reduce the local r.m.s. speed. Two processes that we consider can be important: drag force cooling and collisional cooling. Drag force cooling occurs because part of the kinetic energy exchanged between the particles and the gas is dissipated. Collisional cooling is produced by the highly inelastic collisions between boulders, transferring kinetic energy to heat and deformation. Collisional cooling occurs generally in simulations of resolved collisions in planetary rings²³. In the Supplementary Information we describe how we treat collisional cooling numerically in the self-gravitating simulations by damping the r.m.s. speed of the particles in each grid cell on a collisional timescale. We have found that in the absence of collisional cooling, gravitational collapse still proceeds if the total surface density (of solids and gas) is augmented by 50%. Collisional cooling is thus not a prerequisite of the collapse, but does allow it to occur in somewhat less massive disks. We ignore all other effects of the collisions, such as coagulation and collisional fragmentation. Collisional cooling and self-gravity are turned on after 20 orbits in the self-gravitating simulations.

Our chosen scale-height-to-radius ratio of $H/r = 0.04$ gives a gas temperature of $T = 80$ K at an orbital radius of $r = 5$ AU. For the 256^3 self-gravitating run, we choose the uniform gas volume density to be consistent with the midplane of a disk with surface density of $\Sigma_{\text{gas}} = 300 \text{ g cm}^{-2}$. This corresponds to approximately twice the minimum-mass solar nebula at 5 AU from the (proto-)Sun. An alternative theory for giant planet formation, the disk instability hypothesis^{24,25}, requires column densities at least 20 times higher than the minimum-mass solar nebula for gravitational fragmentation of the gaseous component of the disk to occur.

We have examined the numerical convergence of our models with resolutions ranging from 64^3 to 256^3 zones (see Supplementary Information). The peak particle density on the grid increases with increasing resolution, because of less smoothing in the particle-mesh scheme at higher resolution, resulting in a decrease in the column density threshold for gravitational collapse. Although we have not yet achieved full convergence, our results seem to provide good upper limits to the column density for which collapse can occur. For the self-gravitating simulation, we consider boulders with friction times distributed among $\Omega_K\tau_f = 0.25, 0.50, 0.75, 1.00$. At $r = 5$ AU in our chosen disk model, these correspond to radii of 15–60 cm. Consideration of multiple boulder sizes is vital, as differential aerodynamic behaviour could inhibit gravitational instabilities²⁶. The size range covers roughly half of the two orders of magnitude in particle

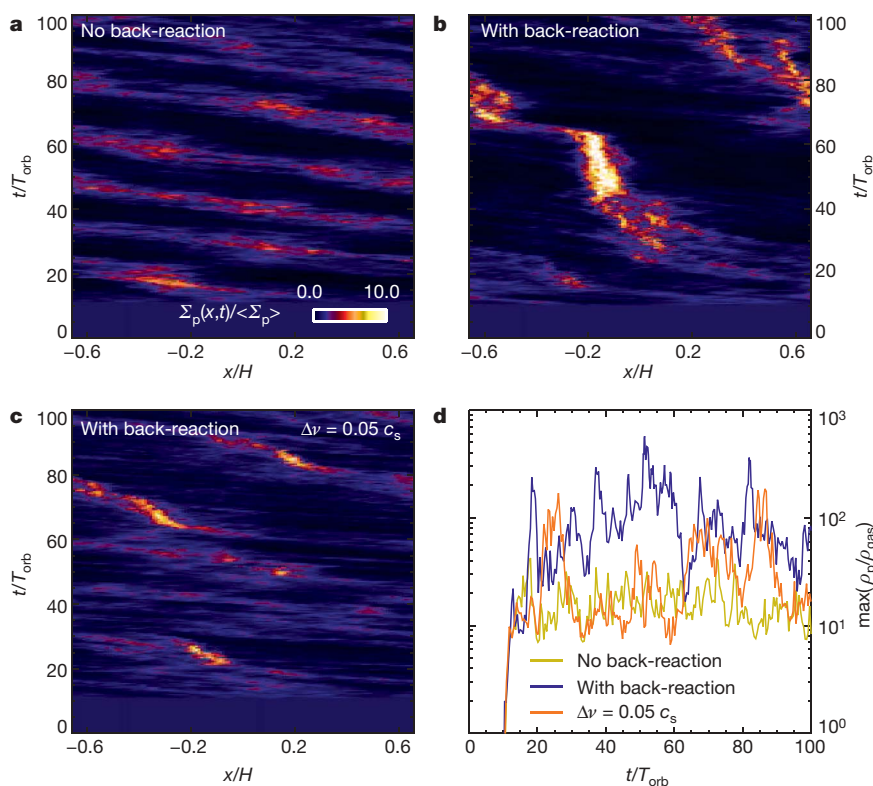


Figure 1 | Topography of the sedimented particle layer in models without self-gravity or collisional cooling. **a**, The azimuthally averaged vertical column density Σ_p of metre-sized boulders (with $\Omega_K\tau_f = 1$) as a function of radial coordinate x (in units of disk scale heights H) and time t (in orbits T_{orb}), in a model where the particles feel gas drag, but the gas does not feel drag from the particles. Radial drift is evident from the tilted bands (particles crossing the inner boundary reappear at the outer). Transient regions of mildly increased gas pressure temporarily concentrate boulders. The gas orbits slightly more slowly on the outer edge of these high pressure regions and slightly faster on the inner edge, resulting in a differential headwind that forces boulders towards the centres of these high-pressure regions^{9,30}. **b**, Including the back-reaction drag force from the particles on the gas allows for the development of the streaming instability, seeded by the existing radial density enhancements. The streaming instability occurs where the collective

drag force of the solids forces the gas to locally move with an orbital speed that is closer to keplerian, reducing the gaseous headwind that otherwise causes boulders to drift radially. Solids then drift into already overdense regions from further out, causing runaway growth in the local bulk density of solids. **c**, The column density when the radial pressure support is changed from $\Delta\nu = -0.02c_s$ to $-0.05c_s$. Radial density enhancements become narrower and shorter-lived owing to downstream erosion of the overdensities by the stronger radial drift. **d**, The maximum particle density ρ_p on the grid (in units of the gas density ρ_{gas}) as a function of time. The average solids-to-gas ratio in the midplane is 0.5, whereas the maximum reaches well over ten times higher values in transient high pressure regions (yellow) and several hundred times higher values when the streaming instability is active (orange and blue).

radius produced by coagulation of microscopic grains²⁷. Smaller particles are ignored, as they are unlikely to separate from the gas and participate in gravitational collapse. In the case of widespread collisional fragmentation, for example, in the warmer terrestrial planet formation region, up to 80% of the solid material may be bound in small fragments^{27,28}, in which case we must implicitly assume an augmentation in solids-to-gas ratio of up to 5 to reach our assumed amount of boulders.

In our self-gravitating model we set $\Delta\nu = -0.02c_s$, but show in the Supplementary Information that gravitationally bound clusters also form for $\Delta\nu = -0.05c_s$, with a factor of two increase in column density threshold. The Supplementary Information also documents that typical boulder collisions happen at speeds below the expected destruction threshold³. We warn, however, that material properties, and thus destruction thresholds, of the boulders are poorly known. Higher resolution studies, and an improved analytical theory of collision speeds that takes into account epicyclic motion, will be needed to determine whether collision speeds have converged; this is because we find an unexplained factor 3 difference for $\Omega_K\tau_f = 1$ particles between typical relative speeds within cells ($\sim 5\text{ m s}^{-1}$) and the expected collision speed of well-mixed particles.

The development of gravitational instability in the 256³ run is shown in Fig. 2. The four different boulder sizes have already accumulated in the same regions (see the block of four joined panels)

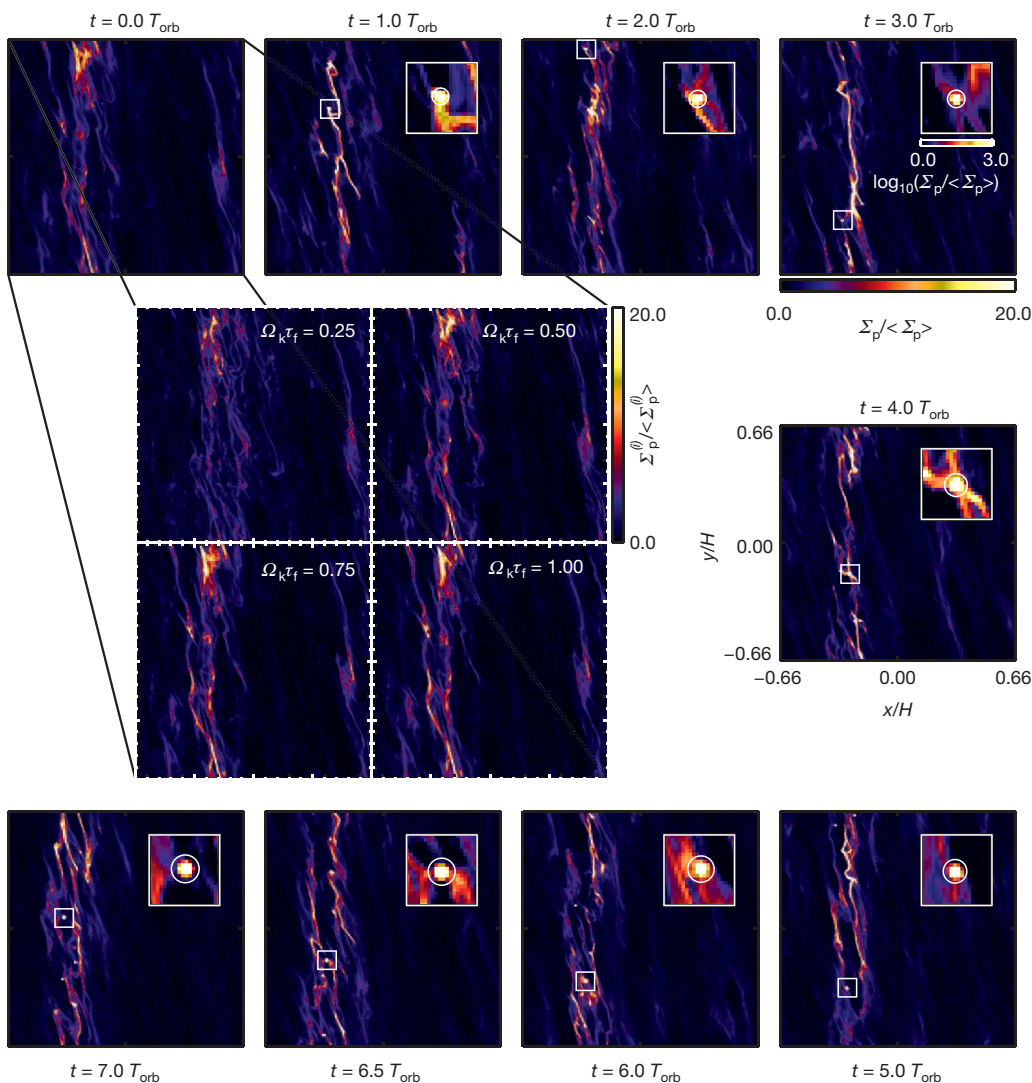


Figure 2 | Time series of the collapse of overdense seeds into gravitationally bound boulder clusters. The central block of four panels shows the column densities of the four different sizes of boulders (in units of the mean column density of each size) plotted independently at a time just before self-gravity is turned on. All four particle sizes have concentrated at similar locations, an important prerequisite for the subsequent gravitational collapse. The surrounding panels show a time series of total column density of solids, in the radial–azimuthal (x – y) plane of the disk, summed over all particle sizes, starting from the upper left and progressing clockwise. Values are normalized to the average value across the grid (see colour bars in upper right panel). Times are given in orbital times T_{orb} after self-gravity is turned on. Inset in each panel is an enlargement of a square region (indicated in the main panel) centred around the Hill sphere of the most massive cluster in the simulation, represented by the white circle. These insets show the log of the column density ratio (see colour bar in upper right panel) to capture the extreme values reached. Overdense bands initially contract radially, forming thin filaments with densities high enough for a full non-axisymmetric collapse into gravitationally bound clumps to take place. As time progresses, the Hill sphere increases in radius as the clusters grow in mass by accreting boulders from the turbulent flow (see Supplementary Video for an animation of this simulation).

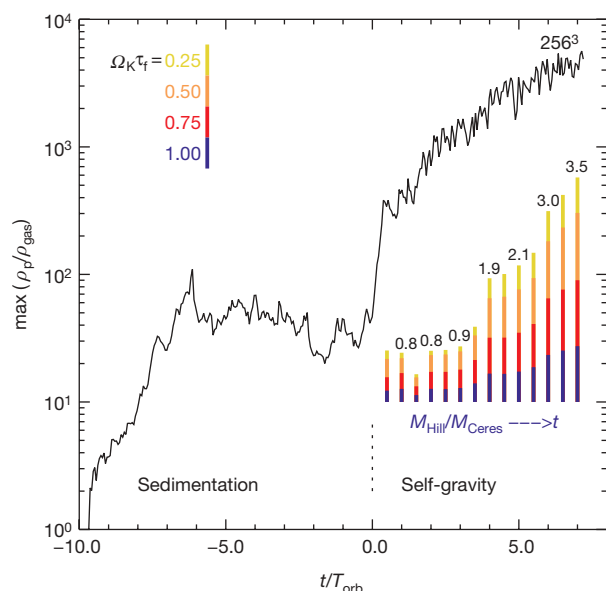


Figure 3 | Mass accretion onto a gravitationally bound cluster. Main plot, the maximum bulk density of solids ρ_p as a function of time, normalized by the average gas density ρ_{gas} . Drag force and vertical gravity are turned on at $t = -10$, whereas self-gravity and collisional cooling are turned on at $t = 0$. The density increases monotonically after the onset of self-gravity because gravitationally bound clusters of boulders form in the mid-plane. After only seven orbits, peak densities in these clusters approach $10^4 \rho_{gas}$ or a million times the average boulder density in the disk. Bottom right inset, coloured bars show (by their height, given on top) the mass contained within the most massive Hill sphere in the box, in units of the mass of the 970-km-radius dwarf planet Ceres ($M_{Ceres} = 9.5 \times 10^{23}$ g). The most massive cluster accretes about $0.5 M_{Ceres}$ per orbit (the entire box contains a total boulder mass of $50 M_{Ceres}$). The cluster consists of approximately equal fractions of the three larger boulder sizes (see colour bar inset at top left). The smallest size, with $\Omega_K \tau_f = 0.25$, is initially underrepresented with a fraction of only 15% because of the stronger aerodynamic coupling of those particles to the gas, but the fraction of small particles increases with time as the cluster grows massive enough to attract smaller particles as well. The mean free path inside the bound clusters is shorter than the size of the cluster, so any fragments formed in catastrophic collisions between the boulders will be swept up by the remaining boulders before being able to escape the cluster (see Supplementary Information).

perhaps be possible owing to radial variation in boulder drift speeds⁶ or photoevaporation of the gas²⁹.) Magnetorotational turbulence thus has a positive effect on the mid-plane layer's ability to gravitationally collapse, although collapse can also occur without it.

The Supplementary Information also includes a model with an adiabatic equation of state and explicit gas heating due to energy dissipated by drag and inelastic collisions. We find that gas heating does not prevent collapse. The maximum temperature reached is not even high enough to melt ice, although that may change with the formation of massive bodies with escape velocity near the sound speed.

Our proposed path to planetesimal formation depends crucially on the existence of a dense sedimentary layer of boulders. Future investigations should focus on the formation and survival of such layers in the light of processes like coagulation, collisional fragmentation and erosion²⁸. Especially important are higher resolution studies of collision speeds, and an improved analytical theory of collisions that includes the epicyclic motion of particles.

Received 19 December 2006; accepted 5 July 2007.

1. Safronov, V. S. *Evolution of the Protoplanetary Cloud and Formation of Earth and the Planets* (NASA Tech. Transl. F-677, Jerusalem, 1972); translation of *Evolutsiia Doplannetnogo Oblaka* (Nauka, Moscow, 1969).

2. Dominik, C., Blum, J., Cuzzi, J. N. & Wurm, G. in *Protostars and Planets V* (eds Reipurth, B., Jewitt, D. & Keil, K.) 783–800 (Univ. Arizona Press, Tucson, 2007).
3. Benz, W. Low velocity collisions and the growth of planetesimals. *Space Sci. Rev.* **92**, 279–294 (2000).
4. Weidenschilling, S. J. Aerodynamics of solid bodies in the solar nebula. *Mon. Not. R. Astron. Soc.* **180**, 57–70 (1977).
5. Goldreich, P. & Ward, W. R. The formation of planetesimals. *Astrophys. J.* **183**, 1051–1062 (1973).
6. Youdin, A. N. & Shu, F. H. Planetesimal formation by gravitational instability. *Astrophys. J.* **580**, 494–505 (2002).
7. Weidenschilling, S. J. & Cuzzi, J. N. in *Protostars and Planets III* (eds Levy, E. H. & Lunine, J. I.) 1031–1060 (Univ. Arizona Press, Tucson, 1993).
8. Hartmann, L. *Accretion Processes in Star Formation* (Cambridge Astrophysics Series No. 32, Cambridge Univ. Press, Cambridge, UK, 1998).
9. Johansen, A., Klahr, H. & Henning, T. Gravoturbulent formation of planetesimals. *Astrophys. J.* **636**, 1121–1134 (2006).
10. Youdin, A. N. & Goodman, J. Streaming instabilities in protoplanetary disks. *Astrophys. J.* **620**, 459–469 (2005).
11. Johansen, A., Henning, T. & Klahr, H. Dust sedimentation and self-sustained Kelvin-Helmholtz turbulence in protoplanetary disk midplanes. *Astrophys. J.* **643**, 1219–1232 (2006).
12. Johansen, A. & Youdin, A. Protoplanetary disk turbulence driven by the streaming instability: Nonlinear saturation and particle concentration. *Astrophys. J.* **662**, 627–641 (2007).
13. Cuzzi, J. N., Dobrovolskis, A. R. & Champney, J. M. Particle-gas dynamics in the midplane of a protoplanetary nebula. *Icarus* **106**, 102–134 (1993).
14. Balbus, S. A. & Hawley, J. F. Instability, turbulence, and enhanced transport in accretion disks. *Rev. Mod. Phys.* **70**, 1–53 (1998).
15. Barge, P. & Sommeria, J. Did planet formation begin inside persistent gaseous vortices? *Astron. Astrophys.* **295**, L1–L4 (1995).
16. Fromang, S. & Nelson, R. P. On the accumulation of solid bodies in global turbulent protoplanetary disc models. *Mon. Not. R. Astron. Soc.* **364**, L81–L85 (2005).
17. Rice, W. K. M., Lodato, G., Pringle, J. E., Armitage, P. J. & Bonnell, I. A. Planetesimal formation via fragmentation in self-gravitating protoplanetary discs. *Mon. Not. R. Astron. Soc.* **372**, L9–L13 (2006).
18. Cuzzi, J. N., Hogan, R. C., Paque, J. M. & Dobrovolskis, A. R. Size-selective concentration of chondrules and other small particles in protoplanetary nebula turbulence. *Astrophys. J.* **546**, 496–508 (2001).
19. Hockney, R. W. & Eastwood, J. W. *Computer Simulation Using Particles* (McGraw-Hill, New York, 1981).
20. Gammie, C. F. Nonlinear outcome of gravitational instability in cooling, gaseous disks. *Astrophys. J.* **553**, 174–183 (2001).
21. Gammie, C. F. Layered accretion in T Tauri disks. *Astrophys. J.* **457**, 355–362 (1996).
22. Tanga, P., Weidenschilling, S. J., Michel, P. & Richardson, D. C. Gravitational instability and clustering in a disk of planetesimals. *Astron. Astrophys.* **427**, 1105–1115 (2004).
23. Salo, H. Gravitational wakes in Saturn's rings. *Nature* **359**, 619–621 (1992).
24. Boss, A. P. Giant planet formation by gravitational instability. *Science* **276**, 1836–1839 (1997).
25. Mayer, L., Quinn, T., Wadsley, J. & Stadel, J. Formation of giant planets by fragmentation of protoplanetary disks. *Science* **298**, 1756–1759 (2002).
26. Weidenschilling, S. J. Can gravitational instability form planetesimals? *Icarus* **116**, 433–435 (1995).
27. Dullemond, C. P. & Dominik, C. Dust coagulation in protoplanetary disks: A rapid depletion of small grains. *Astron. Astrophys.* **434**, 971–986 (2005).
28. Weidenschilling, S. J. The origin of comets in the solar nebula: A unified model. *Icarus* **127**, 290–306 (1997).
29. Throop, H. B. & Bally, J. Can photoevaporation trigger planetesimal formation? *Astrophys. J.* **623**, L149–L152 (2005).
30. Haghighipour, N. & Boss, A. P. On pressure gradients and rapid migration of solids in a nonuniform solar nebula. *Astrophys. J.* **583**, 996–1003 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This collaboration was made possible through the support of the Annette Kade Graduate Student Fellowship Program at the American Museum of Natural History. J.S.O. was supported by the US National Science Foundation, as was M.-M.M.L. in part. We thank J. Cuzzi for discussion about the role of cooling in the gravitational collapse.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.J. (johansen@mpia.de).

LETTERS

The development of a protoplanetary disk from its natal envelope

Dan M. Watson¹, C. J. Bohac¹, C. Hull^{1,2}, William J. Forrest¹, E. Furlan^{3,4}, J. Najita⁵, Nuria Calvet⁶, Paola d'Alessio⁷, Lee Hartmann⁶, B. Sargent¹, Joel D. Green¹, Kyoung Hee Kim¹ & J. R. Houck⁸

Class 0 protostars, the youngest type of young stellar objects, show many signs of rapid development from their initial, spheroidal configurations, and therefore are studied intensively for details of the formation of protoplanetary disks within protostellar envelopes. At millimetre wavelengths, kinematic signatures of collapse have been observed in several such protostars, through observations of molecular lines that probe their outer envelopes. It has been suggested that one or more components of the proto-multiple system NGC 1333–IRAS 4 (refs 1, 2) may display signs of an embedded region that is warmer and denser than the bulk of the envelope^{3,4}. Here we report observations that reveal details of the core on Solar System dimensions. We detect in NGC 1333–IRAS 4B a rich emission spectrum of H₂O, at wavelengths 20–37 μ m, which indicates an origin in extremely dense, warm gas. We can model the emission as infall from a protostellar envelope onto the surface of a deeply embedded, dense disk, and therefore see the development of a protoplanetary disk. This is the only example of mid-infrared water emission from a sample of 30 class 0 objects, perhaps arising from a favourable orientation; alternatively, this may be an early and short-lived stage in the evolution of a protoplanetary disk.

NGC 1333–IRAS 4B (henceforth IRAS 4B) and its neighbour NGC 1333–IRAS 4A (IRAS 4A) lie about 320 pc away⁵. Considered among the archetypal protostars, they are often taken to be a proto-triple system, with IRAS 4B single, and IRAS 4A a 1.8-arcsec binary (projected separation 600 AU), 31 arcsec away from IRAS 4B (ref. 6). Both IRAS 4A and IRAS 4B have high-velocity outflows^{7,8}. That of IRAS 4B is not well resolved spatially, but is presumed to be bipolar, viewed close to the outflow axis^{1,2}. Both also have dense, cold, approximately spheroidal envelopes, which are resolved at millimetre and submillimetre wavelengths. Millimetre-wavelength tracers of dense molecular gas reveal kinematic symptoms of collapse in both envelopes^{1,2}. IRAS 4A is marginally detected in Spitzer Space Telescope–Infrared Array Camera (IRAC) images, and IRAS 4B not at all; only extended emission, resembling scattered light from an outflow cavity viewed close to its axis, is seen by IRAC at 3.6 μ m and 4.5 μ m (R. A. Gutermuth, personal communication).

A decade ago, the ISO Long-wavelength Spectrograph (ISO-LWS) was used to observe the IRAS 4 system^{3,4}. With its 90-arcsec beam, this instrument could not resolve IRAS 4A from IRAS 4B, but it did detect many emission lines thereabouts, among them several low-lying rotational transitions of H₂O. Under the assumption that the water emission arises equally from IRAS 4A and IRAS 4B, observers³ showed that this emission probably probes a cold outer component of the envelopes, and an inner component warm enough ($T > 120$ K) that dust-grain mantles have sublimated to increase greatly the

gas-phase abundance of water. They could not, however, rule out an origin of the water emission in the outflows associated with IRAS 4A and IRAS 4B.

IRAS 4A and IRAS 4B are part of a sample of 30 class 0 objects, which we observed with the Spitzer Infrared Spectrograph (IRS) in 2004–5. Initially we observed these objects in low spectral resolution ($\lambda/\Delta\lambda = 60$ –120), with results as shown in Fig. 1. In many of the surveyed objects we detect spectral lines of molecular hydrogen, and fine-structure lines of low-ionization-potential ions and atoms, usually associated with outflows. One line of this type—[S I] at 25.249 μ m—is the only emission feature that appears in the 20–40- μ m wavelength range for IRAS 4A. The spectrum of IRAS 4B, however, is unique in our sample: it contains many emission features suggestive of spectrally unresolved H₂O lines with a wide range of excitation. We re-observed IRAS 4B in March 2006 with the high-resolution ($\lambda/\Delta\lambda = 600$) IRS module, and confirmed this suggestion

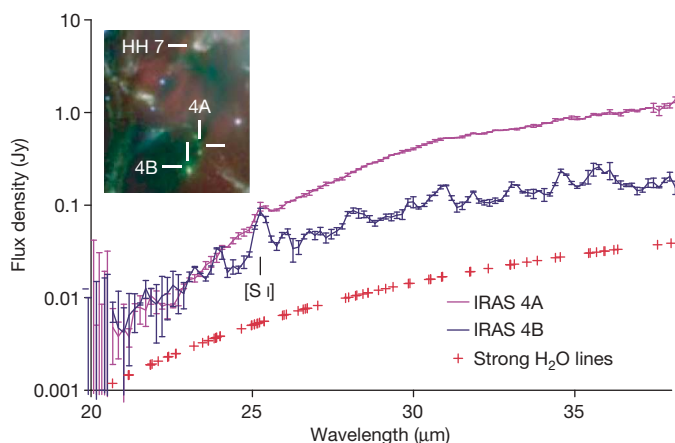


Figure 1 | Spitzer-IRS low-resolution spectra of IRAS 4A and IRAS 4B, and false-colour mid-infrared image. Main panel, spectra of IRAS 4A (magenta trace) and IRAS 4B (blue trace); error bars represent the standard deviations of the flux values in each spectral channel. Wavelengths of strong lines of water, which give rise to blended emission features in the low-resolution spectrum of IRAS 4B, are indicated with red crosses. The IRAS 4 system has total luminosity $28L_{\odot}$ (ref. 19); from these spectra it appears that $4.2L_{\odot}$ belongs to IRAS 4B, and the rest to the two components of IRAS 4A. See Supplementary Information for data-reduction details. Inset, part of the Spitzer-IRAC image-set for NGC 1333 (R. A. Gutermuth, personal communication), with 3.6 μ m shown in blue, 4.6 μ m in green and 8 μ m in red, and positions indicated for IRAS 4A and IRAS 4B, and for the nearby HH 7.

¹Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627-0171, USA. ²Department of Astronomy, University of Virginia, Charlottesville, Virginia 22904, USA. ³NASA Astrobiology Institute, and ⁴Department of Physics and Astronomy, UCLA, Los Angeles, California 90095, USA. ⁵NOAO, Tucson, Arizona 85719, USA. ⁶Department of Astronomy, University of Michigan, Ann Arbor, Michigan 48109, USA. ⁷Centro de Radioastronomía y Astrofísica, UNAM, 58089 Morelia, Michoacán, Mexico. ⁸Center for Radiophysics and Space Research, Cornell University, Ithaca, New York 14853, USA.

in detail (Fig. 2). Besides the [S I] 25.249 μm and [S II] 34.815 μm lines, which can be ascribed to IRAS 4B's outflow, we detect a total of 48 spectral lines that we can assign to 75 pure-rotational transitions of H_2O , seen either singly or in spectrally unresolved combinations. Significantly, all of these lines belong to the ground vibrational manifold of H_2^{16}O ; no vibrationally excited or 'isotopic' water lines are seen. Only the lines at wavelengths greater than 29.8 μm have ever been detected outside the Solar System before, and those only in bright supergiant stars^{9,10}. We also detect seven narrow emission features that we have not succeeded in identifying, but that cannot be due to any well-known molecular or atomic transitions. In the following we will discuss the water-line emission exclusively.

Many important features of the physical state of the emitting material are obtained to good approximation from a simple, conventional model: a plane-parallel slab, with uniform temperature, density high enough that the rotational energy levels of water have thermal-equilibrium populations, and velocity gradient large compared with the ratio of thermal speed and slab thickness. Radiative transfer is included via the escape-probability formalism^{4,11}. We leave the slab's temperature, water column density and emitting area, the extinction towards the slab, and the background continuum temperature, as free parameters. Figure 2 is a plot of the observed spectrum and best-fitting model, and Table 1 a list of model parameters and other quantities we derive from the observations.

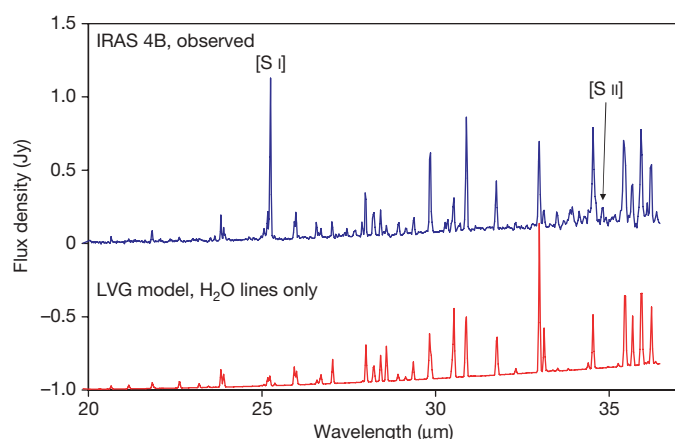


Figure 2 | Comparison of observed and model spectra of IRAS 4B. With the exceptions of the lines of [S I] and [S II] labelled in the plot, the emission lines in the observed spectrum (top) are due to ground-vibrational-state, rotational transitions of H_2^{16}O . (See Supplementary Information for data-reduction and line-identification details.) The model spectrum (bottom) is offset by -1.0 Jy. It is based on a plane-parallel slab, with uniform temperature and density and a large velocity gradient (LVG), treated with the escape-probability formalism^{4,11}. Outside the slab we assume interstellar-like extinction²⁰ from a foreground screen. Parameters of water (energy levels, A -coefficients, and so on) we adopt from the HITRAN database²¹. We suppose that the rotational levels are excited by collisions with hydrogen molecules, and neglect radiative pumping. The latter assumption is justified after modelling, as the line opacities and background intensity are small. However, accurate collisional excitation rate coefficients are only available for two-thirds of the states involved^{22,23}. As a first approximation, we assume thermal-equilibrium populations, and an *ortho*/*para* abundance ratio of 3. The assumption of thermal equilibrium implies a molecular hydrogen density approaching these states' critical densities for collisional de-excitation, which lie in the range 10^{10} – 10^{12} cm^{-3} (see Supplementary Information). The results are not sensitive to the isotopologue relative abundances, for which we adopt the solar abundances of oxygen and hydrogen isotopes. The continuum beneath the lines is blackbody emission to high accuracy, which presumably arises from the optically thick disk behind, or at larger radii than, the slab. This leaves slab temperature, water column density, emitting area, extinction and continuum temperature as free parameters, and we have varied these parameters to produce a minimum- χ^2 fit, with reduced $\chi^2 = 16$ (see Table 1).

Table 1 | Model parameters and inferred quantities for the core of IRAS 4B

| Disk-surface density | Thermal equilibrium, requiring molecular hydrogen density of at least 10^{10} cm^{-3} |
|---|--|
| Disk-surface temperature | 170 K |
| Extinction by envelope | $A_V = 100$ mag, interstellar-like |
| H_2O column density | 9.2×10^{16} cm^{-2} (each face of disk) |
| H_2O -line-emitting mass | 7.5×10^{24} g (H_2O), $\sim 6 \times 10^{27}$ g (total) |
| Emitting area | 6,000 AU^2 (each face of disk) |
| Accretion shock speed | 2 km s^{-1} |
| Total H_2O -line luminosity | $0.03 L_\odot$ (extinction-corrected) |
| Disk accretion rate | $\geq 0.7 \times 10^{-4} M_\odot \text{ yr}^{-1}$ (total) |
| Continuum temperature | 59 K |
| (underlying disk and envelope) | |

Except for the density, these values result from the model fit to the observed spectrum in Fig. 2. Formally, the temperatures are uncertain by ± 2 K, and the extinction is uncertain by ± 50 mag in A_V . The uncertainties in the remaining quantities are dominated by that in the extinction (see Supplementary Information), and are not independent of one another; typically, the maximum uncertainty in these quantities is approximately a factor of two.

The good model fit demonstrates that all the observed rotational states of water are indeed populated close to thermal equilibrium, so the molecular-hydrogen density must approach or exceed these states' critical densities for collisional de-excitation, 10^{10} – 10^{12} cm^{-3} . This is many orders of magnitude too dense to be associated with IRAS 4B's outflow or outer envelope: the source of emission must be in the vicinity of the protostar. But it is not particularly hot at the core of IRAS 4B. The highest-excitation lines we observe are very temperature-sensitive; their faintness relative to the others indicates that there is little material warmer than 170 K.

Predictions from spherical collapsing-envelope models of class 0 objects^{12,13} do not match the observed spectrum; nor do the combinations of density, temperature and emitting area in these models resemble those we have inferred. Our model values are, however, appropriate for dense gaseous disks embedded within such envelopes^{14,15}. Thus the water emission probably shows us, we believe for the first time, accretion from a protostellar envelope by a protoplanetary disk^{14–16}. In Fig. 3 we sketch a cross-section of this arrangement of collapsing envelope and disk. Radiation heating would not suffice to heat such a large area of the disk to 170 K (ref. 17), nor would shocks from spiral density waves; an accretion shock is probably involved. Complex models of our spectrum of IRAS 4B should thus reveal many long-sought details of the assembly of protoplanetary disks.

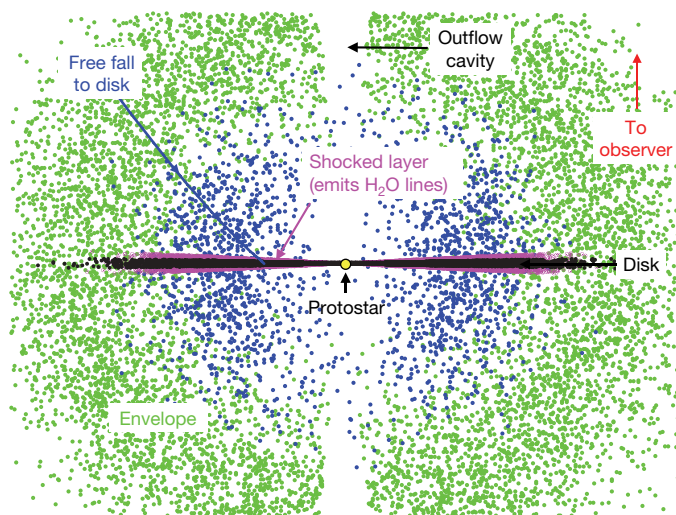


Figure 3 | Cartoon depiction of our model for IRAS 4B. One example streamline (blue line with arrows, top left) is shown to illustrate the infall from the rotating envelope onto the disk¹⁴. The molecular-hydrogen density of the infalling gas (green dots, blue dots) reaches roughly 3×10^9 cm^{-3} or greater (see Supplementary Information) before encountering the disk (black). There the material's deceleration in a 2 km s^{-1} accretion shock produces the warm, extremely dense ($>10^{10}$ cm^{-3}) layer (magenta), with properties as in Table 1, that in turn produces the water-line emission we detect. Our view of the object, as indicated, is close to face-on.

According to the model, the brightest lines have peak optical depth $\tau = 6\text{--}10$, but most of the others are optically thin. From the gas temperature and flux of the optically thick lines, we calculate the emitting area to be $6,000 \text{ AU}^2$ on the side facing us. From the optically thin lines we obtain the mass of water, $7.5 \times 10^{24} \text{ g}$, counting both disk faces. Assuming that all of the solar abundance of oxygen not taken up by CO is present in the gas phase as water ($\text{H}_2\text{O}/\text{H}_2 = 1.5 \times 10^{-4}$)—consistent with rapid sublimation of dust-grain ice mantles expected at $T = 170 \text{ K}$ —the total gas mass is $\sim 6 \times 10^{27} \text{ g}$ ($\sim M_{\text{Earth}}$). The detected water, if condensed, would fill the Earth's oceans about five times.

Also according to the model, the detected water emission comprises about 30% of the total luminosity emitted from water. (The model also predicts far-infrared water lines fainter than those detected by ISO-LWS^{3,4}, suggesting that the latter observations were dominated either by the envelope or the outflow.) Corrected for extinction and this fraction, the water-line luminosity radiated by both faces of the disk is $0.03 L_{\odot}$ (where L_{\odot} is the solar luminosity). Usually^{15,18}, water emission is the dominant coolant of dense molecular material.

Consider material in the envelope, falling freely from a great distance at a steady rate \dot{m} to the disk at radius r , and entering a single, non-magnetic shock at speed v . The post-shock gas temperature T and cooling luminosity L are given by

$$L = \frac{1}{2} \dot{m} v^2 = \frac{GM\dot{m}}{r}, \quad (1)$$

$$T = \frac{2(\gamma-1)\mu}{(\gamma+1)^2 k} v^2,$$

where M is the mass of the protostar, μ the mean molecular mass ($\sim m(\text{H}_2)$), γ the ratio of specific heats ($\sim 7/5$), G the gravitational constant, and k Boltzmann's constant. We have measured T and L , and thence obtain $v = 2 \text{ km s}^{-1}$ and $\dot{m} = 0.7 \times 10^{-4} M_{\odot} \text{ yr}^{-1}$ (where M_{\odot} is the solar mass) similar to that inferred for the outer envelope's infall rate^{1,2}, and much greater than would be inferred for the protostar from its luminosity. We can also evaluate combinations of M and r ; for example, if $r = 50 \text{ AU}$, then $M = 0.14 M_{\odot}$, and the shock extends over a $r = 40\text{--}60 \text{ AU}$ annulus.

At such high density ($10^{10}\text{--}10^{11} \text{ cm}^{-3}$), heating of dust in gas-grain collisions, along with optically thin thermal radiation by this dust, is probably a significant coolant of the post-shock gas. Accordingly, we should regard the mass accretion rate obtained above as a lower limit to the actual rate.

IRAS 4B is alone in our sample of class 0 objects in its mid-infrared water-line emission, and the inferred disk-accretion activity. Perhaps high extinction and unfavourable line-of-sight orientation obscure such emission in the other systems. Alternatively—and more interestingly—we may have observed an early and short-lived stage of protoplanetary disk formation.

Received 2 August 2006; accepted 2 July 2007.

1. Choi, M., Panis, J.-F. & Evans, N. J. II Berkeley-Illinois-Maryland Association Survey of protostellar collapse candidates in HCO^+ and HCN lines. *Astrophys. J. Suppl.* **122**, 519–556 (1999).

2. Di Francesco, J., Myers, P. C., Wilner, D. J., Ohashi, N. & Mardones, D. Infall, outflow, rotation, and turbulent motions of dense gas within NGC 1333 IRAS 4. *Astrophys. J.* **562**, 770–789 (2001).
3. Maret, S., Ceccarelli, C., Caux, E., Tielens, A. G. G. M. & Castets, A. Water emission in NGC 1333 – IRAS 4. *Astron. Astrophys.* **395**, 573–585 (2002).
4. Ceccarelli, C. *et al.* Water line emission in low-mass protostars. *Astron. Astrophys.* **342**, L21–L24 (1999).
5. de Zeeuw, P. T., Hoogerwerf, R., de Bruijne, J. H. J., Brown, A. G. A. & Blaauw, A. A. HIPPARCOS census of the nearby OB associations. *Astron. J.* **117**, 354–399 (1999).
6. Looney, L. W., Mundy, L. G. & Welch, W. J. Unveiling the circumstellar envelope and disk: a subarcsecond survey of circumstellar structures. *Astrophys. J.* **529**, 477–498 (2000).
7. Knee, L. B. G. & Sandell, G. The molecular outflows in NGC 1333. *Astron. Astrophys.* **361**, 671–684 (2000).
8. Blake, G. A. *et al.* A molecular line study of NGC 1333/IRAS 4. *Astrophys. J.* **441**, 689–701 (1995).
9. Neufeld, D. A., Feuchtgruber, H., Harwit, M. & Melnick, G. J. Infrared Space Observatory observations of far-infrared rotational emission lines of water vapour toward the supergiant star VY Canis Majoris. *Astrophys. J.* **517**, L147–L150 (1999).
10. Neufeld, D. A. *et al.* Detection of far-infrared rotational lines of water vapour toward W Hydrae. *Astron. Astrophys.* **315**, L237–L240 (1996).
11. Scoville, N. Z. & Solomon, P. M. Radiative transfer, excitation, and cooling of molecular emission lines. *Astrophys. J.* **187**, L67–L71 (1974).
12. Ceccarelli, C., Hollenbach, D. J. & Tielens, A. G. G. M. Far-infrared line emission from collapsing protostellar envelopes. *Astrophys. J.* **471**, 400–426 (1996).
13. Doty, S. D. & Neufeld, D. A. Models for dense molecular cloud cores. *Astrophys. J.* **489**, 122–142 (1997).
14. Cassen, P. & Moosman, A. On the formation of protostellar disks. *Icarus* **48**, 353–376 (1981).
15. Neufeld, D. A. & Hollenbach, D. J. Dense molecular shocks and accretion onto protostellar disks. *Astrophys. J.* **428**, 170–185 (1994).
16. Ulrich, R. K. An infall model for the T Tauri phenomenon. *Astrophys. J.* **210**, 377–391 (1976).
17. Kenyon, S. J., Calvet, N. & Hartmann, L. The embedded young stars in the Taurus-Auriga molecular cloud. I — Models for spectral energy distributions. *Astrophys. J.* **414**, 676–694 (1993).
18. Neufeld, D. A. & Kaufman, M. J. Radiative cooling of warm molecular gas. *Astrophys. J.* **418**, 263–272 (1993).
19. Sandell, G., Aspin, C., Duncan, W. D., Russell, A. P. G. & Robson, E. I. NGC 1333 IRAS 4: a very young, low-luminosity binary system. *Astrophys. J.* **376**, L17–L20 (1991).
20. Weingartner, J. C. & Draine, B. T. Dust grain-size distributions and extinction in the Milky Way, Large Magellanic Cloud, and Small Magellanic Cloud. *Astrophys. J.* **548**, 296–309 (2001).
21. Rothman, L. S. *et al.* The HITRAN 2004 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transf.* **96**, 139–204 (2004).
22. Green, S., Maluendes, S. & McLean, A. D. Improved collisional excitation rates for interstellar water. *Astrophys. J. Suppl.* **85**, 181–185 (1993).
23. Phillips, T. R., Maluendes, S. & Green, S. Collisional excitation of H_2O by H_2 molecules. *Astrophys. J.* **107**, 467–474 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported in part by NASA through the Spitzer-IRS Instrument Team, Origins and Astrobiology programmes, and by CONACyT (México). We are grateful to R. Gutermuth for use of the Spitzer-IRAC data on NGC 1333, and to M. Jura, L. Keller, G. Sloan and D. Hollenbach for discussions.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.M.W. (dmw@pas.rochester.edu).

Direct observation of second-order atom tunnelling

S. Fölling¹, S. Trotzky¹, P. Cheinet¹, M. Feld¹, R. Saers², A. Widera^{1,3}, T. Müller^{1,4} & I. Bloch¹

Tunnelling of material particles through a classically impenetrable barrier constitutes one of the hallmark effects of quantum physics. When interactions between the particles compete with their mobility through a tunnel junction, intriguing dynamical behaviour can arise because the particles do not tunnel independently. In single-electron or Bloch transistors, for example, the tunnelling of an electron or Cooper pair can be enabled or suppressed by the presence of a second charge carrier due to Coulomb blockade^{1,2}. Here we report direct, time-resolved observations of the correlated tunnelling of two interacting ultracold atoms through a barrier in a double-well potential. For the regime in which the interactions between the atoms are weak and tunnel coupling dominates, individual atoms can tunnel independently, similar to the case of a normal Josephson junction. However, when strong repulsive interactions are present, two atoms located on one side of the barrier cannot separate³, but are observed to tunnel together as a pair in a second-order co-tunnelling process. By recording both the atom position and phase coherence over time, we fully characterize the tunnelling process for a single atom as well as the correlated dynamics of a pair of atoms for weak and strong interactions. In addition, we identify a conditional tunnelling regime in which a single atom can only tunnel in the presence of a second particle, acting as a single atom switch. Such second-order tunnelling events, which are the dominating dynamical effect in the strongly interacting regime, have not been previously observed with ultracold atoms. Similar second-order processes form the basis of superexchange interactions between atoms on neighbouring lattice sites of a periodic potential, a central component of proposals for realizing quantum magnetism^{4–7}.

For the description and observation of quantum mechanical tunnelling, a double-well-type potential, where two localized spatial modes are separated by a barrier, is among the conceptually simplest set-ups. When a particle is initially prepared on one side of this barrier, it will tunnel back and forth between the two sides with a well-defined frequency. For macroscopic quantum systems, such as superconductors or atomic Bose–Einstein condensates, this tunnel coupling can lead to a Josephson-type tunnelling dynamics^{8–10}. When interactions between individual particles are much stronger than the tunnel coupling in the system, quantized Josephson dynamics arises—in which, for example, the charge carriers in superconducting devices tunnel individually across barriers^{11,12}.

In the case of coupled mesoscopic quantum dots, a co-tunnelling regime can be achieved, where separate electrons only tunnel in a correlated way^{13,14}. For ensembles of ultracold atoms in periodic potentials, strong interactions fundamentally alter the properties of the many-body system, leading to strongly correlated phases such as the Mott insulating state^{15–19}. In such cases, where direct first-order tunnelling of single atoms is highly suppressed, second-order correlated tunnelling processes can be the dominant dynamical effects. Despite the absence of direct long-range interaction mechanisms between particles, second-order “superexchange”-type processes

can provide effective spin-dependent interactions between particles at separate positions^{4–7}.

The dynamics of interacting bosonic atoms in a double well with tight confinement is described by a quantized Josephson or a two-mode Bose–Hubbard hamiltonian^{11,12}

$$H = -J(\hat{a}_L^\dagger \hat{a}_R + \hat{a}_R^\dagger \hat{a}_L) - \frac{1}{2} \Delta(\hat{n}_L - \hat{n}_R) + \frac{1}{2} U(\hat{n}_L(\hat{n}_L - 1) + \hat{n}_R(\hat{n}_R - 1)) \quad (1)$$

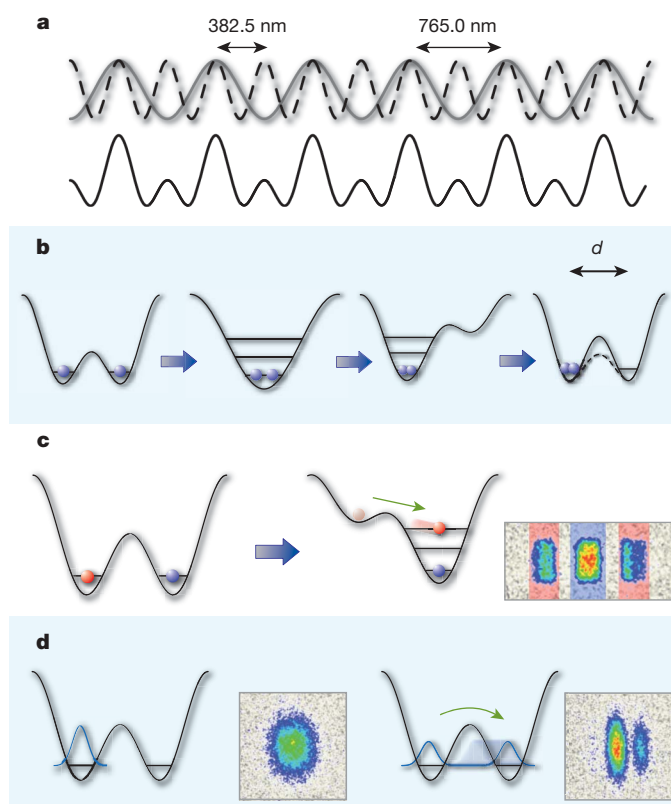


Figure 1 | Schematics of double-well generation, loading and detection sequences. **a**, Superimposing two optical lattice potentials differing in period by a factor of two creates an array of double-well potentials. **b**, Preparation sequence. An initially large well is split into a biased double-well potential such that each left well is populated. The bias is then removed and the central barrier lowered to initiate the tunnelling dynamics (d denotes the well separation). **c**, Position measurement. The atom number on each side can be recorded by ‘dumping’ the population of the left well into an excited vibrational state of the right well²¹. Subsequent band-mapping projects both states into separate Brillouin zones in free space³⁰ (marked red and blue in the inset). **d**, Interferometric detection. After sudden release from the double-well potential and a period of free expansion, the double-slit interference pattern is recorded. Particles localized to one well exhibit no interference; for delocalized atoms the pattern yields the relative single-particle phase ($-\pi/2$ in the case shown).

¹Institut für Physik, Johannes Gutenberg-Universität, 55099 Mainz, Germany. ²Department of Physics, Umeå University, 90187 Umeå, Sweden. ³Institut für Angewandte Physik, Universität Bonn, 53115 Bonn, Germany. ⁴Institute of Quantum Electronics, ETH Zürich, 8093 Zürich, Switzerland.

with J the tunnelling matrix element, $\hat{a}_{L,R}^\dagger$ and $\hat{a}_{L,R}$ the creation and annihilation operators for a bosonic particle in the ground state of the left or right well, Δ the bias potential between the wells and U the interaction energy for two particles in a single well. In the following, we assume repulsive interactions ($U > 0$); however, attractive interactions lead to a similar evolution. The operators \hat{n}_L and \hat{n}_R count the number of atoms in the left and right modes, respectively. In the tight-confinement approximation, the quantum state of the atoms can be described in a Fock state basis $|n_L, n_R\rangle$ with n_L and n_R being non-negative integers. For a single atom, the two possible states $|1, 0\rangle$ and $|0, 1\rangle$ are coupled by the tunnel matrix element J . For the case of two atoms, the states $|2, 0\rangle$ and $|0, 2\rangle$ both directly couple to the state $|1, 1\rangle$ via the tunnelling term in first order. For strong interactions $U \gg J$, the energy difference between the states $|2, 0\rangle$ and $|1, 1\rangle$ is much larger than the coupling, resulting in a strong detuning and therefore in suppressed transitions between these states. For an unbiased junction ($\Delta = 0$), the state $|0, 2\rangle$, however, always has the same energy as $|2, 0\rangle$. The direct transition to this state, which corresponds to the co-tunnelling of both atoms as a pair, is therefore still resonant. This second-order tunnelling process has an effective matrix element $2J^2/U$, which can be obtained by second-order perturbation theory for $J/U \ll 1$ (see refs 5 and 6).

To realize the double-well potentials for ultracold rubidium atoms, we superimpose two periodic potentials with a periodicity of 382.5 nm (short lattice) and 765.0 nm (long lattice) and controllable intensities and relative phase (see Methods). They are produced by two independent optical standing waves in such a way that each potential minimum of the long lattice is split into two wells by a maximum of the short-lattice potential (see Fig. 1a). By changing the relative phase of the two potentials, a controlled bias Δ can be introduced, which can be approximated by the energy difference between the two potential minima. Additional standing waves on the two orthogonal axes provide transverse confinement, creating a three-dimensional array of up to 10^5 double wells²⁰ occupied by one to two ^{87}Rb atoms each. In this array we carry out many identical

instances of the experiment in parallel to obtain the quantum mechanical expectation values of the observables in a single run.

The initial state with the atoms localized on one side of the double wells is obtained by adiabatic changes of the potential after loading in the symmetric configuration. Both sites are combined and the atoms are brought to rest on the left side of their respective double well (see Fig. 1b). The tunnelling dynamics can subsequently be initiated by quickly reducing the barrier height within 200 μs . After an evolution period, we determine the resulting average position and the average single-particle phase relation between the quantum states in the left and right well. The overall sums of atoms N_L and N_R in the left and the right wells, respectively, are obtained by rapidly suppressing tunnelling again in 200 μs and employing a band mapping technique²¹ (see Fig. 1c). We calibrate this method by making reference measurements with fully localized atoms prepared on either side. From the two occupation numbers the average position $\langle x \rangle = (N_R - N_L)/(N_R + N_L)$ of the particles is calculated, which denotes their position relative to the barrier in units of $d/2$, where d is the well separation. The phase relation between the two wells is determined by a separate interferometric sequence, where the lattice is switched off rapidly at the end of the dynamical evolution. The emerging double-slit matter-wave interference pattern then yields the average single-particle phase relation (see Methods and Fig. 1d).

Since approximately 40% of the population is in singly-occupied double wells, we determine the atom pair signal in a two-stage process. First the data point which includes both singly and doubly occupied wells ('total signal') is recorded for a given configuration. The measurement is then repeated with a 'filtering sequence' before the dynamical evolution period, which removes atom pairs from the trap (see Methods). By this, we obtain the single-atom signal, which can then be subtracted from the total signal to obtain the data point for atom pairs.

The time evolution for atom pairs can be calculated using the hamiltonian (1) for any given set of double-well parameters U , J and Δ within a three-state description. Our model for the ensemble

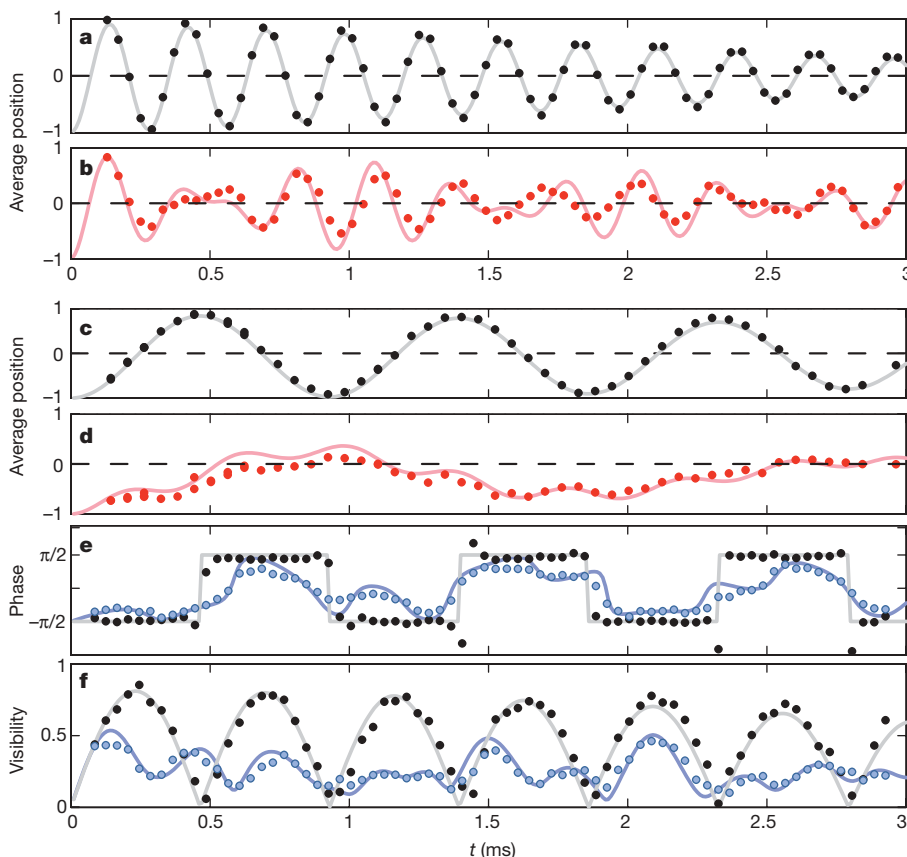


Figure 2 | Tunnelling dynamics. Full dynamical tunnelling evolution for single atoms and atom pairs in the weakly ($J/U = 1.5$) (a, b) and strongly ($J/U = 0.2$) (c–f) interacting regime after initially preparing all particles localized on the left side (position -1) of the double well. The black dots denote the single-atom position (a, c), phase (e) and visibility (f) signal. The red dots show the atom-pair signal (b, d). For strong interactions, first-order tunnelling is suppressed and shifted to the detuned frequency $2.5 \text{ kHz} \approx U/h$. The main dynamical process is pair tunnelling with frequency $4J^2/hU \approx 550 \text{ Hz}$. e, f, Blue dots denote the combined single atom and atom pair signal recorded in the experiment by the interferometric detection method. The solid lines are fits to the data using a model based on the quantized Josephson hamiltonian (see text).

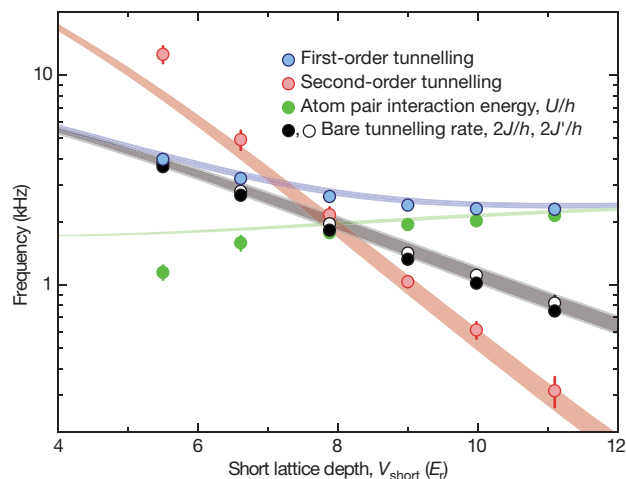


Figure 3 | Tunnelling frequencies versus short-lattice depth (barrier height). The frequencies corresponding to the coupling matrix elements J and J' (see text) are denoted with filled and unfilled black circles, respectively. The crossover above which the measured interaction energy U (green circles) dominates over kinetic energy takes place around $8E_r$. The characteristic frequencies for the second-order tunnelling process and the first-order tunnelling for atom pairs derived from these values are shown as red and blue circles, respectively. Error bars denote the 90% confidence intervals as determined from the fits. The shaded areas show the calculated frequencies for the single-atom tunnelling as well as for the first- and second-order tunnel process as determined from band structure calculations. Their width represents the systematic 2% uncertainty of the lattice depths.

takes into account inhomogeneities in the parameter Δ as well as in the atom density, owing to the overall confining potential (see Methods). We observe experimentally that the parameters in the Bose–Hubbard description in the case of pair occupation are slightly, but notably modified. Specifically, the effective tunnel coupling has to be described by a modified effective single-particle tunnelling rate J' for the two-atom hamiltonian, which is 3–10% higher than the free-particle J . This is in agreement with estimates using perturbative modifications to the wavefunction caused by the interactions.

The measured time-resolved traces resulting from single-atom and atom-pair tunnelling are shown in Fig. 2. The single atom data sets (black dots) show the expected sinusoidal population oscillation

between the two wells at a frequency $2J/h$. If the interaction energy U is smaller than the tunnel matrix element J ($J/U \approx 1.5$; Fig. 2a and b), tunnelling of a single atom out of a pair is only slightly detuned. This process therefore competes with the resonant second-order tunnel process, leading to a signal containing more than one frequency component (see Fig. 2b). When reducing J to reach the interaction-dominated regime ($J/U \approx 0.2$, Fig. 2c–f), the single-atom signal is still sinusoidal at a correspondingly lower frequency. However, the average position of atom pairs now shows a strongly modified behaviour. The pair-breaking first-order process is highly suppressed owing to the detuning by the interaction energy U , and is visible as a small modulation with a period of $\sim 400 \mu\text{s}$ (Fig. 2d). In contrast, the second-order hopping process is now the dominant dynamical effect, leading to a much slower oscillation with a period of $\sim 1.8 \text{ ms}$. Tunnelling is resonant in the centre of the trap; however, in double wells located away from the centre along the double-well axis, the overall confining potential leads to a bias $\Delta \neq 0$. This detuning is especially significant in the case of the second-order process, where the potential bias enters twice and the effective tunnel coupling is low. Thus, for low coupling strengths, the observed oscillation amplitude of the slow second-order process averaged over the ensemble is lower than one, but remains the dominant dynamical process.

In addition to the average centre-of-mass position, the phase and visibility as obtained from the interferometric sequence are shown (Fig. 2e, f). The average direction and velocity of the flow of atoms are characterized by the phase and visibility observables and the tunnelling parameter J . For the single-atom case, we observe distinct jumps in the phase by π , whenever the particle is localized to one side of the potential well and reverses its propagation direction. When atom pairs are included, a much more complex phase evolution emerges. By fitting the modelled dynamical evolution both for atom pairs and single atoms to the data, the tunnelling matrix elements and the interaction energies can be extracted. Figure 3 shows the fitted tunnelling frequency $2J/h$ against the short-lattice depth V_{short} in units of the recoil energy $E_r = \hbar^2/2m\lambda^2$, with m being the mass of the atoms, \hbar Planck's constant and $\lambda = 765 \text{ nm}$ the short-lattice wavelength. We fit both J and J' independently and find good qualitative agreement with the theoretical prediction for J . However, we typically measure a 5% lower coupling than predicted by a band structure calculation, a deviation which is slightly larger than expected from the uncertainties of the lattice depths. The resulting pair tunnelling

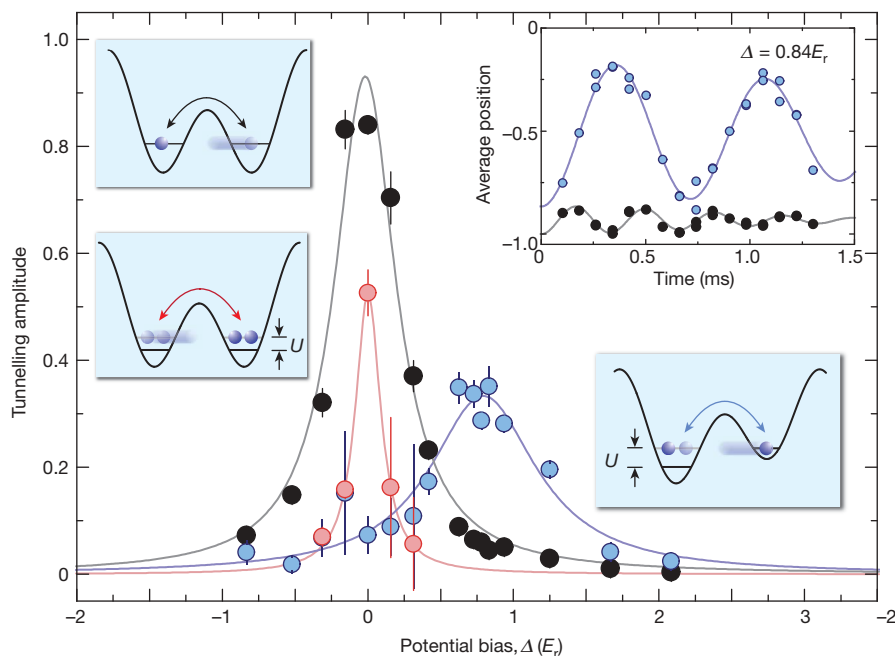


Figure 4 | Conditional tunnelling. The tunnelling amplitude versus the potential bias is measured for the case of single atoms (black data points) and initially doubly occupied lattice sites (blue and red data points) for $V_{\text{short}} = 12E_r$, with $J/U \approx 0.2$. The single-atom tunnelling (depicted in upper left inset) is only resonant in the unbiased case. For doubly occupied sites, two processes corresponding to first- and second-order tunnelling are relevant (lower right and lower left inset) and can be identified by their frequency. Each process shows a distinct resonance. The correlated pair tunnelling (red circles) is resonant for zero bias. Near a potential bias of $\Delta = 0.78(2)E_r$ (centre of lorentzian fitted to blue data points), a conditional tunnelling resonance occurs, where a single atom can tunnel only in the presence of a second atom. The dynamical evolution at this resonance can be seen in the upper right inset, where the blue (black) data points denote the average atom position versus time for the doubly (singly) occupied sites. Without a second atom the tunnelling is strongly suppressed. Amplitudes and error bars are derived from fit parameters and the uncertainties of fitting damped two-component sinusoidal functions to the position signal, as shown in the inset.

frequency $(4J^2/U)/h$ is also plotted, showing the much faster decrease with growing barrier height compared to single-particle tunnelling. For comparison, we plot the interaction strength U as well as the expected frequency of the detuned first-order tunnelling process $\sqrt{4J^2 + U^2}/h$, which asymptotically approaches U as $J \rightarrow 0$.

The interaction-induced suppression of first-order tunnelling for atom pairs is a consequence of energy conservation. Tunnelling can be made resonant again by biasing the single-atom ground state of the well into which the atom tunnels. If the energy offset Δ is equal to U , the first-order process is resonant in the presence of another atom, but detuned by Δ without a second atom. The signal for this conditional tunnelling process is depicted in the inset of Fig. 4, where occupation by two atoms results in a significant amplitude of the sinusoidal tunnelling signal, whereas in the case of a single atom the dynamics is barely visible. The fitted amplitudes of both processes are plotted in Fig. 4 against the bias energy, showing only one resonance at $\Delta = 0$ for single-atom occupation. The two-atom case has two resonances, one for the first-order process at $\Delta = U$ and the other at $\Delta = 0$, where the second-order hopping process is in resonance.

In conclusion, we have reported on the first full characterization of the tunnelling dynamics of an interacting atom pair across a quantum weak link. By tuning the system from a weakly to a strongly interacting regime, we have been able to observe the transition from independent single-particle tunnelling towards correlated tunnelling in second-order exchange-type processes. Our measurement shows that for higher occupations per lattice site, corrections apply to the coupling parameters of the Hubbard model as determined for a single atom. Finally, an additional correlated tunnelling regime was demonstrated, in which a particle can tunnel only on the condition of a second atom being present. Such a single atom switch can be used to efficiently create entanglement over different lattice sites. Starting from spin-triplet pairs in single wells, created for example, via spin-changing collisions²², a resonant tunnelling event can create long-lived entangled singlet or triplet states²³. Super-exchange interactions between neighbouring spin pairs could then be used to engineer large correlated spin-chains, similar to those encountered in cluster states²⁴ or resonating valence-bond-type states^{25,26} of condensed matter physics.

METHODS SUMMARY

The two standing waves are produced by two independent lasers at 765 nm and 1,530 nm, respectively. Their exact wavelength ratio is set by frequency-stabilizing the beat note between the 765 nm light and a frequency-doubled fraction of the 1,530 nm beam.

Control of the relative phase of the periodic potentials at the position of the atoms is possible via the frequency offset of the lock. Dynamically shifting this phase during the state preparation sequence after the initial loading into the lattice makes it possible to prepare the initial state with all atoms on the left side of their respective double well.

The filtering sequence to obtain the pure single-atom tunnelling signal is implemented as a temporary transfer of all atoms to the $F = 2$ hyperfine manifold. During this process, two-body losses due to hyperfine-changing collisions lead to the removal of the atom pairs.

The measured data are modelled by assuming a shell-structure-type distribution of the atoms in the trap to obtain the resulting distribution of double well configurations. The evolution of the total ensemble is calculated by integrating the individual double-well signals over the distribution of bias energies Δ .

To obtain the phase and visibility data, we fit a double-slit interference pattern with gaussian envelope to the time-of-flight images obtained after sudden release of the atoms from the potential. We extract both the relative single-particle phase of the two wells and the visibility of the interference pattern. The modelled time-dependent phase and visibility signal is then fitted to this data to obtain the Hubbard model parameters U , J and Δ .

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 May; accepted 18 July 2007.

1. Averin, D. V. & Likharev, K. K. Coulomb blockade of single-electron tunneling, and coherent oscillations in small tunnel junctions. *J. Low-Temp. Phys.* **62**, 345–373 (1986).
2. Kouwenhoven, L. P. *et al.* in *Mesoscopic Electron Transport* (eds Sohn, L. L., Kouwenhoven, L. P. & Schön, G.) 105–214 (Kluwer, Dordrecht, 1997).
3. Winkler, K. *et al.* Repulsively bound atom pairs in an optical lattice. *Nature* **441**, 853–856 (2006).
4. Auerbach, A. *Interacting Electrons and Quantum Magnetism* (Springer, Berlin, 1998).
5. Duan, L.-M., Demler, E. & Lukin, M. Controlling spin exchange interactions of ultracold atoms in an optical lattice. *Phys. Rev. Lett.* **91**, 090402 (2003).
6. Kuklov, A. & Svistunov, B. Counterflow superfluidity of two-species ultracold atoms in a commensurate optical lattice. *Phys. Rev. Lett.* **90**, 100401 (2003).
7. Altman, E., Hofstetter, W., Demler, E. & Lukin, M. Phase diagram of two-component bosons on an optical lattice. *New J. Phys.* **5**, 113.1–113.19 (2003).
8. Josephson, B. D. Possible new effects in superconductive tunnelling. *Phys. Lett.* **1**, 251–253 (1962).
9. Likharev, K. K. Superconducting weak links. *Rev. Mod. Phys.* **51**, 101–159 (1979).
10. Albiez, M. *et al.* Direct observation of tunneling and nonlinear self-trapping in a single bosonic Josephson junction. *Phys. Rev. Lett.* **95**, 010402 (2005).
11. Averin, D. V. Quantum computing and quantum measurement with mesoscopic Josephson junctions. *Fortschr. Phys.* **48**, 1055–1074 (2000).
12. Makhlin, Y., Schön, G. & Shnirman, A. Quantum-state engineering with Josephson-junction devices. *Rev. Mod. Phys.* **73**, 357–400 (2001).
13. De Franceschi, S. *et al.* Electron cotunneling in a semiconductor quantum dot. *Phys. Rev. Lett.* **86**, 878–881 (2001).
14. Zumbühl, D. M., Marcus, C. M., Hanson, M. P. & Gossard, A. C. Cotunneling spectroscopy in few-electron quantum dots. *Phys. Rev. Lett.* **93**, 256801–256804 (2004).
15. Fisher, M. P. A., Weichman, P. B., Grinstein, G. & Fisher, D. S. Boson localization and the superfluid-insulator transition. *Phys. Rev. B* **40**, 546–570 (1989).
16. Jaksch, D., Bruder, C., Cirac, J. I., Gardiner, C. W. & Zoller, P. Cold bosonic atoms in optical lattices. *Phys. Rev. Lett.* **81**, 3108–3111 (1998).
17. Greiner, M., Mandel, O., Esslinger, T., Hänsch, T. W. & Bloch, I. Quantum phase transition from a superfluid to a Mott insulator in a gas of ultracold atoms. *Nature* **415**, 39–44 (2002).
18. Spielman, I. B., Phillips, W. D. & Porto, J. V. Mott-insulator transition in a two-dimensional atomic Bose gas. *Phys. Rev. Lett.* **98**, 080404 (2007).
19. Stöferle, T., Moritz, H., Schori, C., Köhl, M. & Esslinger, T. Transition from a strongly interacting 1D superfluid to a Mott insulator. *Phys. Rev. Lett.* **92**, 130403 (2004).
20. Anderlini, M., Sebby-Strabley, J., Kruse, J., Porto, J. V. & Phillips, W. D. Controlled atom dynamics in a double-well optical lattice. *J. Phys. B* **39**, S199–S210 (2006).
21. Sebby-Strabley, J. *et al.* Preparing and probing atomic number states with an atom interferometer. *Phys. Rev. Lett.* **98**, 200405 (2007).
22. Widera, A. *et al.* Coherent collisional spin dynamics in optical lattices. *Phys. Rev. Lett.* **95**, 190405 (2005).
23. Roos, C. F. *et al.* Bell states of atoms with ultralong lifetimes and their tomographic state analysis. *Phys. Rev. Lett.* **92**, 220402 (2004).
24. Briegel, H. J. & Raussendorf, R. Persistent entanglement in arrays of interacting particles. *Phys. Rev. Lett.* **86**, 910–913 (2001).
25. Anderson, P. W. The resonating valence bond state in La_2CuO_4 and superconductivity. *Science* **235**, 1196–1198 (1987).
26. Trebst, S., Schollwöck, U., Troyer, M. & Zoller, P. d-Wave resonating valence bond states of fermionic atoms in optical lattices. *Phys. Rev. Lett.* **96**, 250402–250404 (2006).
27. Myatt, C. J., Burt, E. A., Ghrist, R. W., Cornell, E. A. & Wiemann, C. E. Production of two overlapping Bose-Einstein condensates by sympathetic cooling. *Phys. Rev. Lett.* **78**, 586–589 (1997).
28. Schmaljohann, H. *et al.* Dynamics of $F = 2$ spinor Bose-Einstein condensates. *Phys. Rev. Lett.* **92**, 040402 (2004).
29. Fölling, S., Widera, A., Müller, T., Gerbier, F. & Bloch, I. Formation of spatial shell structures in the superfluid to Mott insulator transition. *Phys. Rev. Lett.* **97**, 060403 (2006).
30. Greiner, M., Bloch, I., Mandel, O., Hänsch, T. W. & Esslinger, T. Exploring phase coherence in a 2D lattice of Bose-Einstein condensates. *Phys. Rev. Lett.* **87**, 160405 (2001).

Acknowledgements We thank A. M. Rey and B. Paredes for discussions. We acknowledge funding through the DFG and the European Union (MC-EXT QUASICOMBS). R.S. acknowledges support by the EU QUEDIS programme as well as SJCKMS and the Kempe I and II foundations.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to I.B. (bloch@uni-mainz.de).

METHODS

Double-well lattice potential. The light for the two standing waves along the double-well axis is created by a 1,530 nm fibre laser and a Ti:sapphire laser running at 765 nm. The frequency relation is controlled by frequency-doubling a small part of the 1,530 nm output and recording the beat note between this light and the 765 nm lattice light. This beat note is stabilized to any offset frequency within a 1 GHz interval by a feedback circuit that adjusts the frequency of the fibre laser using a piezo-mechanical actuator. In our set-up, both laser beams are retro-reflected by a common mirror, such that the relative phase of the two periodic potentials is fixed at the mirror position. The phase slip over the distance between mirror and atoms (25 cm) caused by choosing a non-zero offset frequency allows us to control the relative phase of the two standing waves at the position of the atoms, thereby choosing the well geometry and bias energy. Typical potential depths are $9.5E_r$ for the long lattice and $33E_r$ for each 843 nm transversal lattice beam, in units of the 765 nm recoil energy E_r , resulting in a parabolic confinement of $\sim 2\pi \times 80$ Hz in the longitudinal direction.

State preparation. After creating an almost pure BEC of typically 8×10^4 atoms, we ramp up the optical potential within 160 ms. During the ramp, inter-double-well tunnelling is suppressed, but not the tunnelling through the barrier ($10E_r$ final short-lattice depth). This results in a majority of the atoms being loaded into doubly-occupied sites. We then adiabatically remove the central barrier to bring the atom pairs together in one well (Fig. 1b). This well is moved to the left by half of the short-lattice period before the short lattice is ramped up again. By subsequently moving the long-wavelength lattice back to the original position we arrive again at a symmetric potential with all atoms now on the left side of their respective double well.

Filtering sequence. To obtain the tunnelling signal for single atoms, we transfer all atoms from the $F = 1$, $m_F = -1$ hyperfine state to the $F = 2$, $m_F = 0$ hyperfine state by a microwave adiabatic rapid passage after merging the wells. This strongly enhances spin relaxation collisions^{22,27,28}, which release enough energy to efficiently remove both atoms of each pair from the lattice. After a 40 ms hold time, the remaining atoms in singly occupied sites are transferred back to the $F = 1$, $m_F = -1$ state.

Model. To predict the time evolution of an individual double-well system, we diagonalize the hamiltonian (equation (1)) for the single-atom and the two-atom cases independently and calculate the time evolution for the two initial states $|1,0\rangle$ and $|2,0\rangle$. Owing to the gaussian shape of all lattice beams, the tunnelling, interaction and bias parameters are not perfectly identical for all double wells of the array. We assume a shell-structure distribution²⁹ of the atoms, with an outer region of singly occupied double wells and a spherical inner core of doubly occupied wells. This corresponds to a zero-temperature assumption. For our chosen lattice ramps and double well configuration we do not expect a

perfect shell structure, but the obtained dynamical evolution is not very sensitive to changes of the distribution shape. For example, using a thermal distribution of the atoms with temperatures comparable to U/k_B and $J = 0$ gives similar results. Here, k_B denotes the Boltzmann constant. We only take into account the inhomogeneity of λ across the cloud due to the harmonic confinement. The resulting dephasing of the ensemble is the only cause of damping of the total signal within our model. Other effects such as the inhomogeneity of the tunnelling matrix element as well as all inhomogeneities due to imperfections of the beam shapes are not included. We assume a fixed outer radius of 30 sites and use the trap frequency of the confinement and the radius of the inner shell as fitting parameters. When fitting the model to the phase and visibility data from measurements, additional sources of damping can be present. These cause an over-estimation of the trapping frequency obtained from the fit, which we observe to be up to 200%. The effect is most prominent towards the large J/U regime, where the simple three-mode approximation starts to fail. The observed damping of the position signal (Fig. 2a, b) in this regime cannot be reproduced with our model and is included by an additional, empirically determined exponential damping term with a decay time of 3.5 ms.

Phase measurement and fitting. To determine the phase relation between the wells, the trap is switched off and the interference pattern from the wavefunctions emerging from the two wells is recorded. The image is integrated perpendicularly to the double-well axis and a double-slit interference pattern with gaussian envelope is fitted to the resulting profile:

$$P(x) = A e^{-(x-x_0)^2/W^2} [1 + V \cos(kx + \theta)] \quad (2)$$

The fitting parameters of the envelope are the amplitude A , the width W and average centre x_0 . The interference pattern parameters are the visibility V , the periodicity k as well as the phase θ , which is directly given by the average single-particle phase relation between the wells. The momentum-space Wannier-function envelope of the measured profiles is not perfectly gaussian, especially for large tunnel coupling, leading to an overall decrease in visibility. For fitting the tunnelling dynamics model, the time-dependent signal obtained in this way is used. We simultaneously fit the signal for the total (mixed) population and for the filtered population of singly occupied double wells as shown in Fig. 2. We do not perform the atom-number-sensitive separation of the interference patterns into the signals from singly and doubly occupied sites, and model the mixed signal directly for fitting. Apart from the traces shown in Fig. 2, we typically do not record the average position for fitting, because the interference analysis yields more information per single image than the determination of the average position and the resulting data allows the determination of all parameters shown in Fig. 3.

Increased subaerial volcanism and the rise of atmospheric oxygen 2.5 billion years ago

Lee R. Kump¹ & Mark E. Barley²

The hypothesis that the establishment of a permanently oxygenated atmosphere at the Archaean–Proterozoic transition (~2.5 billion years ago) occurred when oxygen-producing cyanobacteria evolved¹ is contradicted by biomarker evidence for their presence in rocks 200 million years older². To sustain vanishingly low oxygen levels despite near-modern rates of oxygen production from ~2.7–2.5 billion years ago thus requires that oxygen sinks must have been much larger than they are now. Here we propose that the rise of atmospheric oxygen occurred because the predominant sink for oxygen in the Archaean era—enhanced submarine volcanism—was abruptly and permanently diminished during the Archaean–Proterozoic transition. Observations^{3–5} are consistent with the corollary that subaerial volcanism only became widespread after a major tectonic episode of continental stabilization at the beginning of the Proterozoic. Submarine volcanoes are more reducing than subaerial volcanoes⁶, so a shift from predominantly submarine to a mix of subaerial and submarine volcanism more similar to that observed today would have reduced the overall sink for oxygen and led to the rise of atmospheric oxygen.

Abundant evidence supports the argument that the atmosphere before ~2.5 billion years ago (~2.5 Gyr ago) was essentially devoid of oxygen⁷ (Fig. 1). The strongest argument comes from the presence of sedimentary sulphides with mass-independent fractionation (MIF) only in rocks older than 2.45 Gyr (ref. 8), a phenomenon that requires both the virtual absence ($<10^{-5}$ times the present atmospheric level) of molecular oxygen⁹ and an abundance of reducing gases (such as methane¹⁰) in the atmosphere. The recent recognition of 2.92- and 2.76-Gyr-old sedimentary sulphides without a strong MIF sulphur isotope signature^{11,12} supports an earlier suggestion¹³ that atmospheric oxygen levels may have increased temporarily before the permanent rise 2.45 Gyr ago, although a collapse of atmospheric

methane without a corresponding rise of atmospheric O₂ cannot be ruled out¹⁰.

The lack of any secular trend in the carbon isotopic composition of the ocean, as preserved in limestones, argues against a substantial increase in organic carbon burial (the long-term source of oxygen) between 2.5 and 2.4 Gyr ago⁶ (but see ref. 14 for an opposing view). Thus, most explanations focus on a decrease in consumption rate^{6,15–17}. The absence of oxidized soil profiles and red beds indicates that oxidative weathering rates were negligible during the Archaean. The likely sinks for oxygen were its reaction with reduced volcanic and metamorphic gases¹⁵ and its reaction with Fe²⁺, either during hydrothermal alteration of seafloor basalts, or as hydrothermal Fe²⁺ dissolved in anoxic sea water¹⁸.

For volcanic sinks to have been larger in the Archaean than in the post-Archaean, volcanic/metamorphic gases must have been more reducing: even though rates of volcanism were higher in the Archaean, the surficial redox balance would only be tipped towards reducing if these gases were more reduced because, as the carbon isotopes show, the flux of oxygen through organic matter burial scaled up proportionately⁶.

The coincidence of the rise of atmospheric O₂ with the Archaean–Proterozoic transition suggests that the stabilization of continental cratons may have affected the volcanic sink for oxygen. Although there is evidence for the existence of ‘continental’ crust since the Eoarchaean (since 3.6 Gyr ago)⁵, the oxygen isotope ratios of igneous zircons¹⁹ and trace-element and isotopic compositions of sedimentary rocks³ are not consistent with the presence of large, elevated, mature, continental landmasses before 3.0 Gyr ago. The majority of Archaean cratons achieved long-term tectonic stability as elevated continental land masses between 2.7 and 2.5 Gyr ago with buoyant, depleted upper-mantle keels²⁰ and thermo-mechanically stable

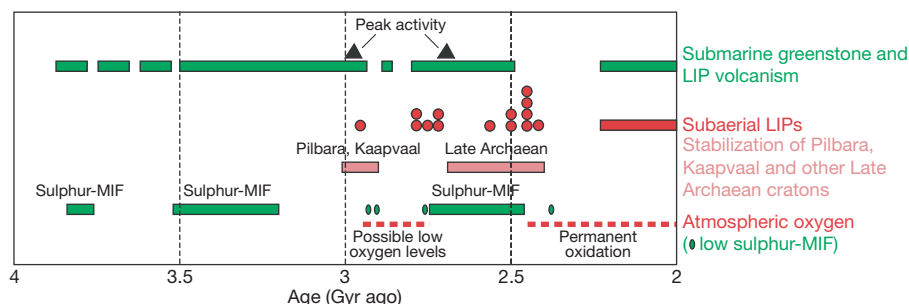


Figure 1 | Archaean–Palaeoproterozoic volcanism, continental stabilization, and atmospheric oxygen evolution. The bars labelled ‘Sulphur-MIF’ represent evidence of strong MIF of sulphur isotopes. This strong constraint on atmospheric oxygen levels is recorded in three sedimentary sequences 3.85–3.77, 3.5–3.2 and 2.74–2.45 Gyr ago⁸. We note

the abrupt and permanent end of the strong MIF signal after 2.45 Gyr ago. The oval symbols indicate well-dated sedimentary sequences with low sulphur-MIF values^{8,11,12}, consistent with low and rising levels of atmospheric oxygen. This figure is based on Supplementary Tables 1 and 2.

¹NASA Astrobiology Institute and Department of Geosciences, Pennsylvania State University, 535 Deike Building, University Park, Pennsylvania 16802, USA. ²School of Earth and Geographical Sciences, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia.

crustal profiles²¹. This occurred during and after a period of mantle plume activity (2.72–2.66 Gyr ago) that resulted in the most prodigious episode of generation and preservation of continental crust in Earth history (Fig. 1). The Pilbara and Kaapvaal cratons achieved tectonic stability during an earlier period of mantle plume activity between 3.0 and 2.9 Gyr ago. Archaean cratons amalgamated to form the first supercontinent or a number of supercratons by the end of the Archaean^{13,22,23}.

Archaean volcanism seems to have been dominated by submarine eruptions, with greenstone belts characterized by submarine basalt–komatiite successions and dominantly submarine island–arc-like volcanic complexes²⁴. Many of the preserved basalt–komatiite successions are mantle–plume-derived large igneous provinces (LIPs; preserved as flood basalts and large mafic dyke swarms). The arc-like successions contain only minor reworked pyroclastic deposits derived from subaerial eruptions, and subaerial andesite volcanoes on thickened continental crust seem to be almost absent in the Archaean. Various explanations for the abundance of Archaean submarine volcanic rocks have been proposed, but most depend on the idea that the Archaean oceanic lithosphere would have been thicker and that isostatic adjustment displaced the oceans onto the continents²⁵.

In contrast, the preserved Palaeoproterozoic volcanic record is dominated by subaerial continental LIPs²⁴. Most Proterozoic greenstone successions resemble Phanerozoic island arc assemblages with a higher proportion of rocks derived from subaerial eruptions than their Archaean counterparts²⁴.

It is possible that the predominance of immature submarine crust in preserved Archaean cratons is an artefact of selective preservation: only a small fraction of the crust that existed in the Archaean is preserved today. However, if plate tectonics and tectonic recycling operated in a steady-state mode with a near-modern balance of large stable continents and oceans in the Archaean, then one might expect that cratons would be preferentially preserved and immature subduction/accretion terrains would be preferentially destroyed³, rather than the reverse. Immature tectonic environments were probably the norm in the Archaean, with fragments of evolving proto-continental crust within largely submarine volcanic arcs at convergent plate margins. Because such proto-continental fragments

would take several hundred million years to achieve thermo-mechanical stability^{5,21}, they would have been tectonically weak and recycled at a similar rate to volcanic arcs and oceanic plateaus.

In summary, although it is possible that the predominance of juvenile crust in the Archaean is a preservational artefact²⁶, there is considerable evidence to support the opposing view that the transition from an Archaean to modern-day plate-tectonic regime occurred in two major steps, in the Mesoarchaean and at the Archaean–Proterozoic boundary, and that this transformation promoted subaerial volcanism.

A database of LIPs through time²⁷ and 24 greenstone successions older than 2.0 Gyr provides a further indication of the shift in the balance between submarine and subaerial volcanism (Figs 1 and 2; Supplementary Tables 1 and 2). Archaean greenstone successions contain both submarine continental and oceanic LIPs as well as numerous submarine arc-like volcanic successions. Most pre-2.5-Gyr-ago (Archaean) LIPs and arc-like successions were emplaced during periods of increased mantle plume activity at ~2.95 and ~2.7 Gyr ago, with seven subaerial Archaean flood basalt sequences, ranging in age from 2.94 to 2.71 Gyr old, restricted to the Pilbara and Kaapvaal cratons (which functioned as stable cratons by ~2.9 Gyr ago). In contrast, of the 28 Palaeoproterozoic 2.5–2.0-Gyr-old LIPs listed²⁷, 25 (89%) are subaerial. Seven dominantly subaerial continental LIPs were emplaced on four cratons between 2.51 and 2.41 Gyr ago, during a period of increased mantle plume activity^{4,13}. A subsequent period of increased magmatic activity between 2.24 and 2.0 Gyr ago is recorded by two submarine LIPs, four submarine greenstone successions and 20 subaerial continental LIPs worldwide.

We interpret these data to be consistent with our proposal that the stabilization of Archaean cratons (to form the first large tectonically stable high-standing continents) resulted in a change from dominantly submarine volcanism to a mixture of submarine and subaerial volcanism in tectonic settings and proportions similar to the Phanerozoic Earth. Interestingly, stabilization of the Pilbara and Kaapvaal cratons corresponds to the early occurrence of low-MIF sedimentary sulphides; the Mozaan group in the Kaapvaal craton¹² overlies the 2.94-Gyr-old Pongola subaerial LIP and is similar in age to the Mosquito Creek group in the Pilbara¹¹, and the Hardy formation in the Pilbara¹¹ falls between the 2.78-Gyr-old Mount Roe and 2.74-Gyr-old Kylenea subaerial LIPs (Fig. 1). However, the most profound and permanent change occurred abruptly, coinciding with the ~2.51- to 2.45-Gyr-long period of increased mantle plume activity and subaerial volcanism during the final stages of amalgamation of Late Archaean cratons to form the first large stable continents or supercontinent⁴.

The redox state of volcanic gases differs markedly between subaerial and submarine eruptions^{6,17}. Volcanic gases that erupted subaerially have generally equilibrated at high temperatures and low pressures with magmas close to the fayalite–magnetite–quartz buffer. Consequently, oxidized gases (H₂O, CO₂ and SO₂) dominate. Submarine volcanoes erupt at lower temperatures and thus are more reducing¹⁷. Hydrothermal vent fluids equilibrate with a suite of alteration minerals at intermediate temperatures and higher pressure, and are thus also more reduced fluids with higher concentrations of H₂, CO, CH₄ and H₂S.

Holland⁶ characterizes the reducing power of volcanic gases with a term f defined as:

$$f = \frac{m_{\text{H}_2} + 0.6m_{\text{CO}} + 3.6m_{\text{CH}_4} + 3m_{\text{H}_2\text{S}} - 0.4m_{\text{CO}_2}}{3.5(m_{\text{SO}_2} + m_{\text{H}_2\text{S}})} + \frac{1}{3.5}$$

modified by us to include CH₄. Here m_i represents the molal concentration of species i in hydrothermal fluids or the mole or volume fraction of i for volcanic gases, and the numbers are derived from the stoichiometries of the reactions that produce organic matter and pyrite from the volcanic gases. f values greater than 1, if representative of the global average, imply that there is more than enough reducing power to convert 20% of the CO₂ to organic C (based on the C

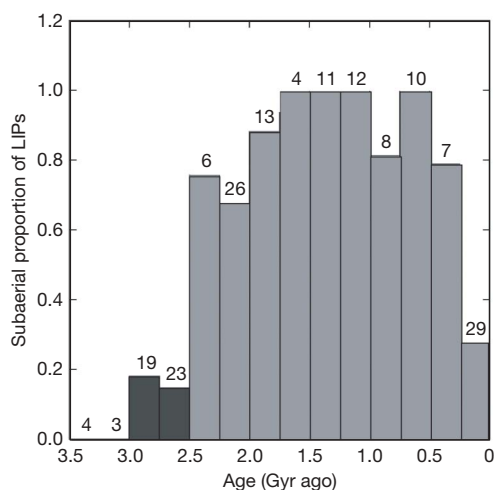


Figure 2 | Secular variation in proportion of subaerial LIPs. Proportion determined as the percentage of the total subaerial LIP occurrences in the age range divided by the sum of the percentage of the total occurrences of both subaerial and submarine LIPs in that age range (see Supplementary Tables 1 and 2). The numbers indicate the total number of LIPs reported. Dark bars reflect Archaean occurrences (before the rise of atmospheric oxygen); light bars reflect post-Archaean occurrences after the rise of atmospheric oxygen. The low proportion over the last 250 million years is probably a preservational bias; no sea floor older than 180 Myr has been preserved.

isotope constraint) and all of the sulphur to pyrite, allowing H_2 to accumulate in the atmosphere. Alternatively, one may think of Holland's f as indicative of the reducing power of volcanic gases, which when combined with the total flux of volcanic gases gives the oxygen demand of volcanic emanations. Oxygen levels rise when the demand for oxygen falls below its supply.

The non-flux-weighted average f value for subaerial volcanic gases (including methane) is (mean \pm standard error) 0.48 ± 0.14 (see Supplementary Tables 3 and 4 for data used and statistical analysis). The range is quite large, from f values below -2 (that is, insufficiently reducing to support 20% conversion of CO_2 to organic matter) to values greater than $+4$. Conceivably, the flux-weighted average could lie towards one end of this range if more- or less-oxidizing volcanoes dominated the global flux. However, most of the values for modern subaerial volcanoes fall between 0.2 and 0.7 (Fig. 3). In contrast, the range of f values for hydrothermal fluids is much smaller, and the average f of 1.04 ± 0.04 falls well above the range for most subaerial volcanoes. A two-tailed Student's t -test assuming unequal variances confirms that the sample means are different, with the probability that the means are the same being 0.0003. With the caveat that our estimate of the average subaerial f value is subject to revision based on a more complete accounting of subaerial volcanic fluxes and their redox state, the relative proportions of subaerial and submarine volcanism today cited by Holland⁶ give a global f value of ~ 0.69 . This value falls well below the threshold for anoxia ($f=1$), as one would expect given the oxygenated state of today's atmosphere.

The data and arguments presented above support the conclusion that the proportion of subaerial volcanism in the Archaean was low. This alone may have been just sufficient to create an anoxic atmosphere, given that submarine volcanoes have an f value very near 1. However, in a sulphate-free Archaean ocean²⁸, the H_2 concentration of hydrothermal fluids would have been up to two orders of magnitude larger²⁹, even though the H_2S concentration would have been similar to average vent fluids today. Using the same CO_2 concentration as Holland's average⁶ (2.0×10^{-2} molal) and an H_2 concentration of 1.8×10^{-2} m (figure 1 in ref. 29) produces an f value for typical hydrothermal volcanism of the Archaean of 1.5 (CH_4 not included). In other words, there would have been at least 50% excess reducing power in Archaean submarine volcanic emissions beyond

that used by organisms for reducing carbon dioxide and consumed by the reduction of volcanic SO_2 . This excess hydrogen would have been available to accumulate in the atmosphere or be used by methanogens^{16,30}; in either case, a reducing atmosphere free of molecular oxygen would have existed.

We thus propose that the rise of atmospheric oxygen was closely tied to Earth's tectonic evolution (Fig. 1). Evidence for a possible early rise to low oxygen levels between ~ 2.92 and 2.74 Gyr ago coincides with stabilization of the Pilbara and Kaapvaal cratons and subsequent continental subaerial flood basalt volcanism, with the most intense episode of dominantly submarine mantle plume activity recorded in the Earth's history occurring 2.72 – 2.66 Gyr ago and resulting in a return to anoxic conditions. The abrupt switch at ~ 2.5 Gyr ago from dominantly submarine volcanic eruptions to a mixture of submarine and subaerial eruptions similar to that of the Phanerozoic was coincident with strong evidence for the permanent establishment of atmospheric oxygen. Either cyanobacterial evolution was somehow tied to the tectonic event 2.5 Gyr ago, or more likely, cyanobacterial oxygenic photosynthesis evolved earlier, but an overwhelming volcanic sink delayed its expression in the atmosphere. As oxygen levels rose, the sink for oxygen associated with submarine volcanism was replaced by the modern weathering sink, including the weathering of subaerial volcanoes, and these sinks grew with atmospheric oxygen levels until a new atmospheric steady state was achieved that was sufficiently oxygen-rich to quench the preservation of MIF sulphur isotope effects.

Received 21 February; accepted 29 June 2007.

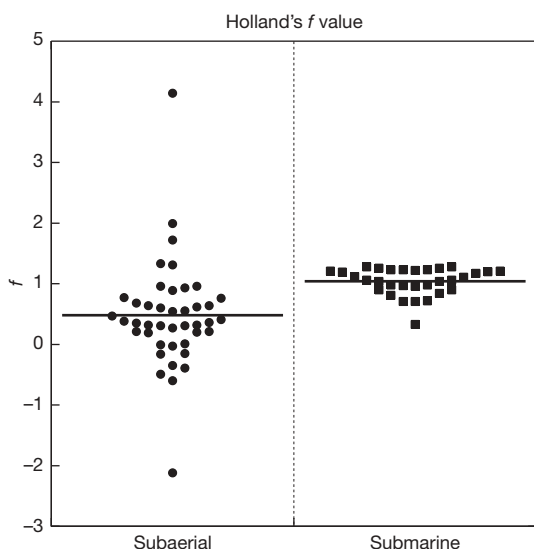


Figure 3 | The distribution of Holland's f values⁶ for modern submarine and subaerial volcanoes. The calculation includes CH_4 . See Supplementary Table 3 for documentation of sources. A global average of $f > 1$ is sufficiently reducing to allow H_2 accumulation in the atmosphere. $f = 1$ reduces 20% of CO_2 to organic matter and all the SO_2 to pyrite. $f = 0$ reduces 20% of CO_2 to organic carbon but no SO_2 to pyrite. $f < 0$ has insufficient reducing power to reduce 20% of CO_2 to organic carbon.

- Kopp, R. E., Kirschvink, J. L., Hillburn, I. A. & Nash, C. Z. The Paleoproterozoic snowball Earth: A climate disaster triggered by the evolution of oxygenic photosynthesis. *Proc. Natl Acad. Sci. USA* **102**, 11131–11136 (2005).
- Brocks, J. J., Logan, G. A., Buick, R. & Summons, R. Archaean molecular fossils and the early rise of eukaryotes. *Science* **285**, 1033–1036 (1999).
- Veizer, J. & Mackenzie, F. T. in *Treatise on Geochemistry* (eds Holland, H. D. and Turekian, K. K.) 369–407 (Elsevier, Amsterdam, 2004).
- Barley, M. E., Bekker, A. & Krapez, B. Late Archaean to Early Paleoproterozoic global tectonics, environmental change and the rise of atmospheric oxygen. *Earth Planet. Sci. Lett.* **238**, 156–171 (2005).
- Hawkesworth, C. J. & Kemp, A. I. S. Evolution of the continental crust. *Nature* **443**, 811–817 (2006).
- Holland, H. D. Volcanic gases, black smokers, and the great oxidation event. *Geochim. Cosmochim. Acta* **66**, 3811–3826 (2002).
- Canfield, D. E. The early history of atmospheric oxygen: Homage to Robert M. Garrels. *Annu. Rev. Earth Planet. Sci.* **33**, 1–36 (2005).
- Papineau, D., Mojzsis, S. J. & Schmitt, A. K. Multiple sulfur isotopes from Paleoproterozoic Huronian interglacial sediments and the rise of atmospheric oxygen. *Earth Planet. Sci. Lett.* **255**, 188–212 (2007).
- Pavlov, A. A. & Kasting, J. F. Mass-independent fractionation of sulfur isotopes in Archaean sediments: strong evidence for an anoxic Archaean atmosphere. *Astrobiology* **2**, 27–41 (2002).
- Zahnle, K. J., Claire, M. W. & Catling, D. C. The loss of mass-independent fractionation in sulfur due to a Palaeoproterozoic collapse of atmospheric methane. *Geobiology* **4**, 271–283 (2006).
- Ohmoto, H., Watanabe, Y., Ikemi, H., Poulson, S. R. & Taylor, B. E. Sulphur isotope evidence for an oxic Archaean atmosphere. *Nature* **442**, 908–911 (2006).
- Ono, S., Beukes, N. J., Rumble, D. & Fogel, M. L. Early evolution of atmospheric oxygen from multiple-sulfur and carbon isotope records of the 2.9 Ga Mozaan group of the Pongola supergroup, southern Africa. *South Afr. J. Geol.* **109**, 97–108 (2006).
- Barley, M. E., Krapez, B., Groves, D. I. & Kerrich, R. The Late Archaean bonanza: metallogenic and environmental consequences of the interaction between mantle plumes, lithospheric tectonics and global cyclicity. *Precamb. Res.* **91**, 65–90 (1998).
- Bjerrum, C. J. & Canfield, D. E. New insights into the burial history of organic carbon on the Early Earth. *Geochem. Geophys. Geosyst.* **5**, doi:10.1029/2004GC000713 (2004).
- Kump, L. R., Barley, M. E. & Kasting, J. F. Rise of atmospheric oxygen and the "upside-down" Archaean mantle. *Geochem. Geophys. Geosyst.* **2**, doi:10.1029/2000GC000114 (2001).
- Claire, M. W., Catling, D. C. & Zahnle, K. J. Biogeochemical modelling of the rise in atmospheric oxygen. *Geobiology* **4**, 239–269 (2006).
- Li, Z.-X. A. & Lee, C.-T. A. The constancy of upper mantle fO_2 through time inferred from V/Sc ratios in basalts. *Earth Planet. Sci. Lett.* **228**, 483–493 (2004).
- Hayes, J. M. & Waldbauer, J. R. The carbon cycle and associated redox processes through time. *Phil. Trans. R. Soc. Lond. B* **361**, 931–950 (2006).

19. Valley, J. W. *et al.* 4.4 billion years of crustal maturation: oxygen isotope ratios of magmatic zircon. *Contrib. Mineral. Petrol.* **150**, 561–580 (2005).
20. Sleep, N. H. Evolution of the continental lithosphere. *Annu. Rev. Earth Planet. Sci.* **33**, 369–393 (2005).
21. Sandiford, M. & McLaren, S. in *Evolution and Differentiation of the Continental Crust* (eds Brown, M. & Rushmer, T.) 67–91 (Cambridge Univ. Press, Cambridge, 2005).
22. Bleeker, W. The late Archean record: A puzzle in ca. 35 pieces. *Lithos* **71**, 99–134 (2003).
23. Condie, K. C. Episodic continental growth and supercontinents: a mantle avalanche connection? *Earth Planet. Sci. Lett.* **163**, 97–108 (1998).
24. Condie, K. C. in *Archean Crustal Evolution* (ed. Condie, K. C.) 85–120 (Elsevier, Amsterdam, 1994).
25. Moores, E. M. Pre-1 Ga (pre-Rodinian) ophiolites: Their tectonic and environmental implications. *Geol. Soc. Am. Bull.* **114**, 80–95 (2002).
26. Hoffman, P. F. Geological constraints on the origin of the mantle root beneath the Canadian shield. *Phil. Trans. R. Soc. Lond. A* **551**, 523–532 (1990).
27. Prokoph, A., Ernst, R. E. & Buchan, K. L. Time-series analysis of large igneous provinces: 3500 Ma to present. *J. Geol.* **112**, 1–22 (2004).
28. Habicht, K. S., Gade, M., Thamdrup, B., Berg, P. & Canfield, D. E. Calibration of sulfate levels in the Archean ocean. *Science* **298**, 2372–2374 (2002).
29. Kump, L. R. & Seyfried, W. E. Hydrothermal Fe fluxes during the Precambrian: Effect of low oceanic sulfate concentrations and low hydrostatic pressure on the composition of black smokers. *Earth Planet. Sci. Lett.* **235**, 654–662 (2005).
30. Kharecha, P., Kasting, J. & Siefert, J. A coupled atmosphere-ecosystem model of the early Archean Earth. *Geobiology* **3**, 53–76 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements L.R.K. acknowledges support from the NASA Astrobiology Institute and the National Science Foundation. M.E.B.'s contribution was supported by the Australian Research Council.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.R.K. (lkump@psu.edu).

Projected increase in continental runoff due to plant responses to increasing carbon dioxide

Richard A. Betts¹, Olivier Boucher¹, Matthew Collins¹, Peter M. Cox^{1,2}, Peter D. Falloon¹, Nicola Gedney³, Deborah L. Hemming¹, Chris Huntingford⁴, Chris D. Jones¹, David M. H. Sexton¹ & Mark J. Webb¹

In addition to influencing climatic conditions directly through radiative forcing, increasing carbon dioxide concentration influences the climate system through its effects on plant physiology¹. Plant stomata generally open less widely under increased carbon dioxide concentration², which reduces transpiration^{1,3–6} and thus leaves more water at the land surface⁷. This driver of change in the climate system, which we term ‘physiological forcing’, has been detected in observational records of increasing average continental runoff over the twentieth century⁸. Here we use an ensemble of experiments with a global climate model that includes a vegetation component to assess the contribution of physiological forcing to future changes in continental runoff, in the context of uncertainties in future precipitation. We find that the physiological effect of doubled carbon dioxide concentrations on plant transpiration increases simulated global mean runoff by 6 per cent relative to pre-industrial levels; an increase that is comparable to that simulated in response to radiatively forced climate change (11 ± 6 per cent). Assessments of the effect of increasing carbon dioxide concentrations on the hydrological cycle that only consider radiative forcing^{9–11} will therefore tend to underestimate future increases in runoff and overestimate decreases. This suggests that freshwater resources may be less limited than previously assumed under scenarios of future global warming, although there is still an increased risk of drought. Moreover, our results highlight that the practice of assessing the climate-forcing potential of all greenhouse gases in terms of their radiative forcing potential relative to carbon dioxide does not accurately reflect the relative effects of different greenhouse gases on freshwater resources.

Physiological forcing of the hydrological cycle by CO₂ has been shown to be the major cause of historical increases in continental-scale runoff⁸. However, predictions of future runoff change due to anthropogenic climate change generally do not consider this^{9–11}, partly because uncertainty in precipitation change is considered to be the main limitation⁹.

We performed a perturbed-physics ensemble of 224 doubled-CO₂ experiments with the HadSM3 climate model, which is a mixed-layer ocean version of the HadCM3 general circulation model^{12,13} but including the MOSES land surface scheme^{4,8}. The simulated runoff depends on precipitation (both amount and intensity) and evapotranspiration (the sum of evaporation and transpiration), and the latter also affects climate through the surface energy and moisture budgets⁴. Transpiration depends on canopy conductance, the large-scale aggregate of stomatal conductance that responds to temperature, humidity, soil moisture and photosynthetically active radiation, and CO₂ concentration^{4,14–17}.

Each of the 224 experiments used a different version of HadSM3. The versions differed in the values assigned to certain key model

parameters¹², with multiple parameters perturbed concurrently¹³. Each experiment consisted of a pair of simulations using the same model version, one simulation with pre-industrial CO₂ and one with doubled CO₂. Such techniques allow the variation of results between ensemble members to give some indication of uncertainty in the predicted climate response^{12,18}.

In our ensemble, one perturbation was the choice of whether to include physiological responses to the CO₂ increase. All 224 ensemble members included radiative forcing due to CO₂, but 81 members also included physiological forcing while the remaining 143 members did not. These two sub-ensembles are labelled RADPHYS and RAD respectively. The number of members of RADPHYS and RAD was an arbitrary consequence of the procedure for selecting members of the 224-member ensemble for general purposes, and not a specific selection for this work alone. We were therefore able to extend the ensemble technique to examine the effect of a particular parameter, in this case the switch for the inclusion of physiological forcing, in the context of uncertainties arising from variations in the other parameters (see Methods).

In our analysis, Y_1 and Y_2 represent 20-year area mean runoff for the pre-industrial (subscript 1) and doubled-CO₂ (subscript 2) members of an individual pair of simulations, ΔY is the difference within a pair, and \bar{Y}_1 , \bar{Y}_2 and $\Delta \bar{Y}$ represent the means over a sub-ensemble. Similar notation holds for precipitation P , with area means taken over land only, and for r , which is the ratio of Y to P :

$$Y = rP \quad (1)$$

The sub-ensemble mean global runoff in the control simulations (\bar{Y}_1) was approximately 2% higher in RADPHYS than RAD (Table 1), but the standard deviations in Y_1 were an order of magnitude larger than this difference and a t -test showed the difference in \bar{Y}_1 to be not statistically significant. Global runoff increased with doubling CO₂ in all pairs of simulations in both RADPHYS and RAD; the ensemble-mean increase $\Delta \bar{Y}$ was $43 \pm 15 \text{ kg m}^{-2} \text{ yr}^{-1}$ in RADPHYS, but only $27 \pm 11 \text{ kg m}^{-2} \text{ yr}^{-1}$ in RAD (Table 1, Fig. 1a). This difference in $\Delta \bar{Y}$ was significant at the 0.1% level. The doubled-CO₂ runoff increase was therefore approximately 59% larger when physiological forcing is included in this set of simulations.

Global P and \bar{P} also increased with doubling CO₂ in both RADPHYS and RAD (Table 1, Fig. 1b), and the sub-ensemble mean $\Delta \bar{P}$ was greater in RADPHYS. Although this might appear inconsistent with a relative reduction in the return of moisture to the atmosphere because of relatively decreased transpiration, it is consistent with the enhanced warming seen over land (Fig. 1c) arising from the reduced evaporative cooling^{1,3,4}. The reduced recycling of moisture over land was offset by an increase in moisture convergence from over the oceans.

¹Met Office Hadley Centre, Fitzroy Road, Exeter EX1 3PB, UK. ²School of Engineering, Computing and Mathematics, Exeter University, Exeter EX4 4QF, UK. ³Met Office, Joint Centre for Hydro-Meteorological Research, ⁴Centre for Ecology and Hydrology, Maclean Building, Wallingford, Oxfordshire, OX10 8BB, UK.

Table 1 | Physiological forcing effects on global mean precipitation, runoff and runoff/precipitation ratio

| | RAD | RADPHYS | RADPHYS – RAD | RAD_DV | RADPHYS_DV | RADPHYS_DV – RAD_DV |
|---------------------------------------|-------------|-------------|---------------|--------|------------|---------------------|
| \bar{Y}_1 | 252 ± 54 | 257 ± 57 | 5 | 220 | 220 | 0 |
| \bar{Y}_2 | 279 ± 49 | 300 ± 56 | 21 | 227 | 238 | 11 |
| $\Delta\bar{Y}$ | 27 ± 11 | 43 ± 15 | 16 | 7 | 18 | 11 |
| \bar{P}_1 | 682 ± 102 | 731 ± 104 | 49 | 706 | 702 | –4 |
| \bar{P}_2 | 718 ± 108 | 769 ± 109 | 51 | 706 | 691 | –15 |
| $\Delta\bar{P}$ | 36 ± 16 | 38 ± 19 | 2 | 0 | –11 | –11 |
| \bar{r}_1 | 0.37 ± 0.05 | 0.35 ± 0.05 | –0.02 | 0.31 | 0.31 | 0 |
| \bar{r}_2 | 0.39 ± 0.05 | 0.39 ± 0.05 | 0.00 | 0.32 | 0.34 | 0.02 |
| $\Delta\bar{r}$ | 0.02 ± 0.01 | 0.04 ± 0.01 | 0.02 | 0.01 | 0.03 | 0.02 |
| $\frac{\Delta\bar{Y}}{\bar{Y}_1}$ (%) | 11 ± 6 | 17 ± 5 | 6 | 3 | 8 | 5 |
| $\frac{\Delta\bar{P}}{\bar{P}_1}$ (%) | 5 ± 2 | 5 ± 3 | 0 | 0 | –2 | –2 |
| $\frac{\Delta\bar{r}}{\bar{r}_1}$ (%) | 6 ± 4 | 12 ± 4 | 6 | 3 | 10 | 7 |

Y and P are given in units of $\text{kg m}^{-2} \text{yr}^{-1}$. Standard deviations refer to variations in the long-term means of experiments within a sub-ensemble. RAD_DV and RADPHYS_DV results are 30-year means for 2000–2030 (subscript 1) and 2070–2100 (subscript 2) in one pair of transient simulations with dynamic vegetation, with no standard deviations because only one experiment was performed.

However, $\Delta\bar{P}$ was only 6% greater in RADPHYS than RAD (significant at 5%). For a given global ΔP , the global ΔY was generally higher in RADPHYS than in RAD (Fig. 1d). \bar{r} increased by 0.02 on doubling CO_2 in RAD (Table 1), possibly as a result of the warmer climate featuring more intense precipitation that exceeded the infiltration rate of the soil more often than at present-day levels of CO_2 . However, \bar{r} increased by 0.04 on doubling CO_2 in RADPHYS, consistent with a reduction in transpiration.

To quantify the relative contribution of changes in precipitation and evapotranspiration to the runoff changes, we approximate the total change in runoff Y in terms of separate contributions from changes in P and r , as a percentage of the baseline runoff:

$$\frac{\Delta Y}{Y_1} = \frac{\Delta P}{P_1} + \frac{\Delta r}{r_1}$$

(2)

This linear approximation is found to be valid for Y and P in both RAD and RADPHYS (Table 1). The differences in the means of $\Delta Y/Y_1$, $\Delta P/P_1$ and $\Delta r/r_1$ between RADPHYS and RAD are approximately 6%, 0 and 6% respectively (Table 1). This demonstrates that increases in r are the dominant cause of the greater increase in Y in RADPHYS than RAD.

At continental scales (Fig. 2 and Supplementary Information), where P increased with doubling CO_2 such as in Asia (Fig. 2b, Supplementary Table 2), North America (Fig. 2d, Supplementary Table 4), and in most ensemble members in Europe (Fig. 2c, Supplementary Table 3) and Oceania (Fig. 2e, Supplementary Table 5), Y increased more in RADPHYS than in RAD. Where P decreased, such as in most ensemble members in South America (Fig. 2f, Supplementary Table 6) and in a large number of ensemble members in Africa (Fig. 2a, Supplementary Table 1), Y decreased less or increased more in RADPHYS than in RAD. Indeed in Europe, \bar{P}

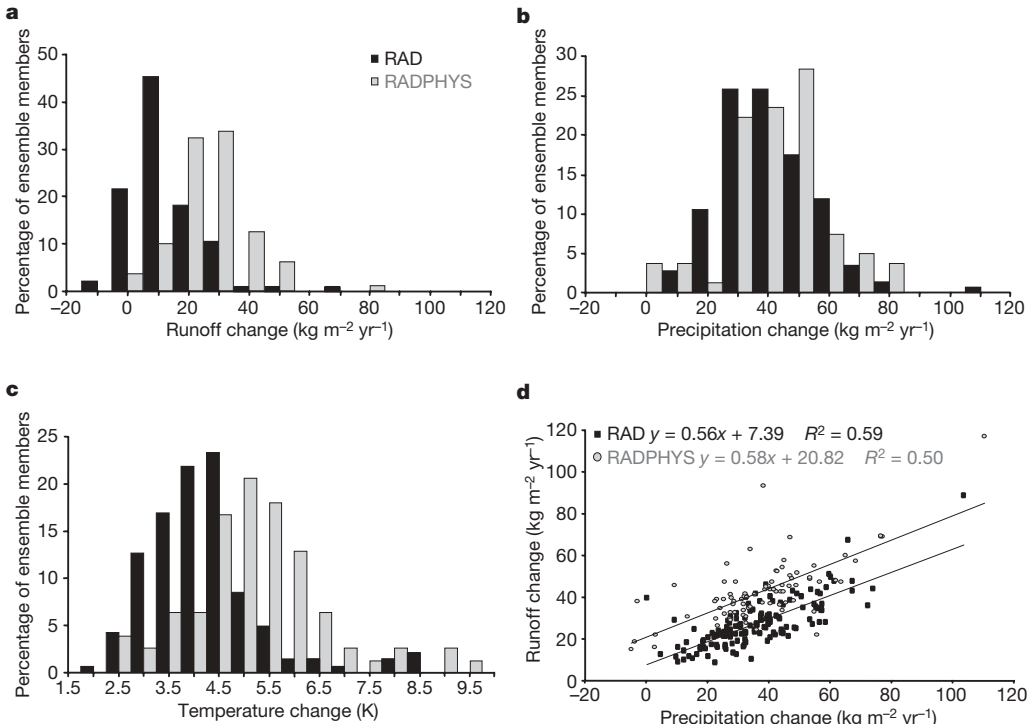


Figure 1 | Impact of physiological forcing on global mean runoff, precipitation and temperature. a, Frequency distribution of simulated changes in global mean runoff due to doubling CO_2 in a 224-member perturbed-physics global climate model ensemble, with physiological forcing included (RADPHYS) and excluded (RAD). **b**, As for **a** but for

precipitation over land. **c**, As for **a** but for near-surface air temperature over land. **d**, Global mean runoff changes versus global land mean precipitation changes due to doubling CO_2 in 224 ensemble members, with physiological forcing included and excluded.

increased less in RADPHYS than RAD, whereas \bar{Y} increased more in RADPHYS (Supplementary Table 3). Similarly in Africa, \bar{P} decreased more in RADPHYS than RAD, whereas \bar{Y} increased in RADPHYS but decreased in RAD (Supplementary Table 1). In all cases, \bar{r} increased more in RADPHYS than RAD at the continental scale as well as the global scale.

A further aspect of physiological forcing is CO_2 fertilization of photosynthesis, which can affect changes in leaf area index and vegetation distribution^{3,18,19}. Increased leaf area index can act to increase canopy conductance and potentially offset stomatal closure, and changes in leaf area index and vegetation type can also affect climate through changes in land surface properties such as albedo and aerodynamic roughness^{3,20,22,23}. Changes in leaf area index and vegetation distribution were not included in RAD and RADPHYS. Previous studies^{18,21} showed increased Y with physiological forcing for a number of models that included dynamic vegetation and variable leaf area index, but these did not include feedbacks to the atmosphere, which may bias the result. To provide a more complete assessment directly comparable with the current work, we performed two additional simulations RAD_DV and RADPHYS_DV including dynamic vegetation²⁴, variable leaf area index and an ocean general circulation model²⁵. In RAD_DV, CO_2 acted only as a greenhouse gas, whereas in RADPHYS_DV, CO_2 changes affected stomatal closure and also fertilized photosynthesis. These were transient simulations to account for vegetation dynamics timescales, driven by the IS92a CO_2 concentration scenario²⁶ in which CO_2 approximately

doubles at the end of the twenty-first century compared to the present day. In both simulations, large-scale vegetation dynamics were included, but in RAD_DV the vegetation responded only to climate change, whereas in RADPHYS_DV the vegetation also responded to physiological forcing through both stomatal responses and fertilization of photosynthesis.

Both RADPHYS_DV and RAD_DV simulated increasing Y as CO_2 increased, but RADPHYS_DV showed a more rapid increase (Table 1). RAD_DV simulated very little change in global land mean P , despite an increase in overall global (land+ocean) mean precipitation, as a result of significant decreases in Amazonia and some other regions offsetting increases elsewhere. RADPHYS_DV simulated a decrease in P , largely because the decrease over Amazonia was greater than in RAD. This was partly a result of reduced transpirational return of moisture to the atmosphere, and partly a result of a northward shift in the Inter-Tropical Convergence Zone attributed to increased Northern Hemisphere warming due to a net reduction in canopy conductance and decreased surface albedo arising from increased leaf area¹⁹. However, despite this difference in ΔP being opposite in sign to the difference in $\Delta \bar{P}$ between the RAD and RADPHYS ensembles that excluded dynamic vegetation, the difference in $\Delta r/r_i$ between RADPHYS_DV and RAD_DV was 7%, which is similar to the 6% difference between RADPHYS and RAD. This suggests that the influence of physiological forcing on r is not significantly modified by changes in leaf area index or vegetation distribution, at least at the global scale. Nevertheless, we note the potential

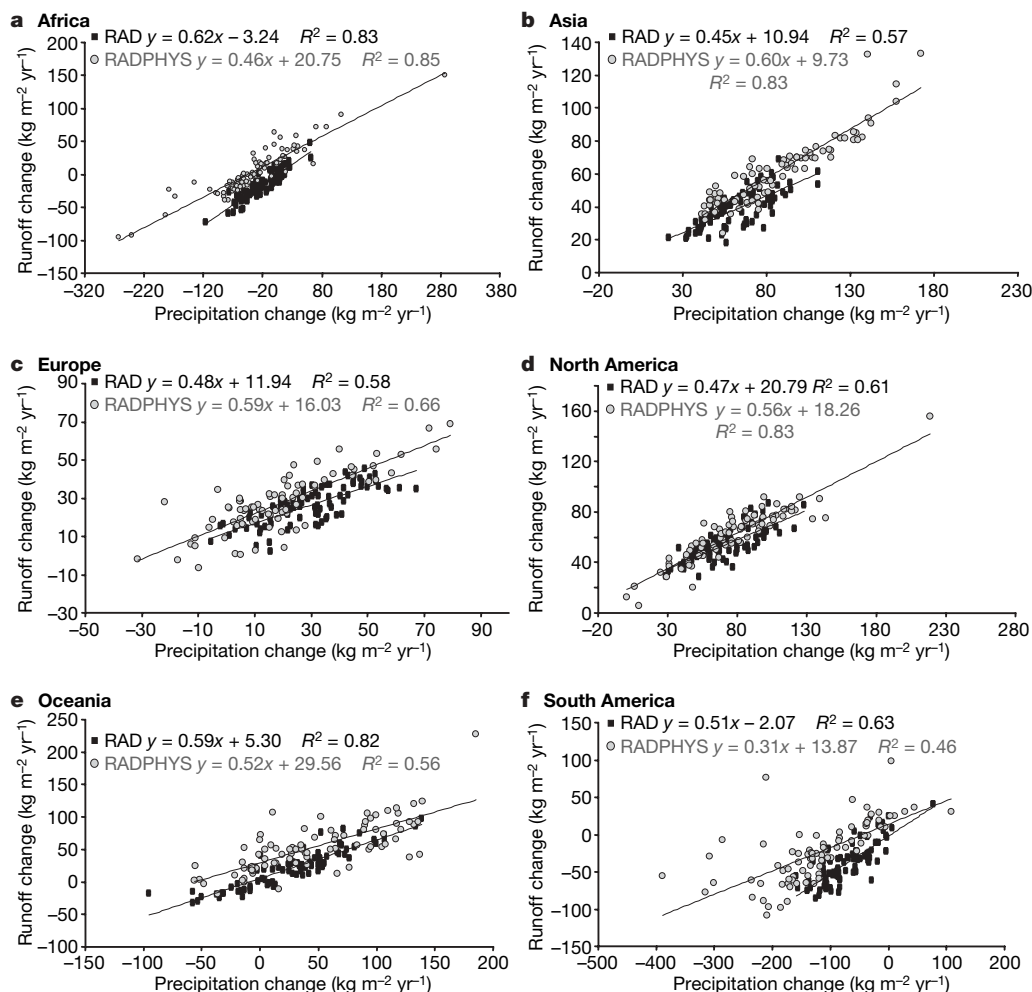


Figure 2 | Impact of physiological forcing on relationship between changes in runoff and precipitation on doubling CO_2 . Continental mean runoff changes versus precipitation changes due to doubling CO_2 in 224 ensemble

members, with physiological forcing included (RADPHYS) and excluded (RAD). **a**, Africa; **b**, Asia; **c**, Europe; **d**, North America; **e**, Oceania; **f**, South America.

for large impacts of vegetation dynamics on ΔY at regional scales, through both the local effects of transpiration changes and the remote effects of precipitation change induced by vegetation changes elsewhere.

Because the effect of physiological forcing on doubled- CO_2 runoff changes is important even in the context of uncertainties in precipitation change, assessments of climate change impacts on future flood and drought risk should take this into account. With greater increases and smaller decreases in runoff due to physiological forcing, the risks of rain and river flooding may increase more than has previously been anticipated, because intense precipitation events would be more likely to occur over saturated ground. In contrast, the risks of hydrological drought may not increase as much as expected on the basis of meteorological changes alone. However, reduced precipitation is not completely negated by physiological forcing, so some regions may still experience increased drought.

The strong influence of physiological forcing on runoff also raises an important issue regarding the comparison of CO_2 with other greenhouse gases. The United Nations Framework Convention on Climate Change (UNFCCC)²⁷ requires that concentrations of different greenhouse gases are routinely compared in terms of a 'CO₂ equivalent', which for non- CO_2 greenhouse gases is conceptualized as the concentration of CO_2 which would exert the same influence on climate. The UNFCCC and the Kyoto Protocol currently quantify this with global warming potentials (GWPs) based on radiative forcing²⁸.

However, this assumes that radiative forcing is the only mechanism through which greenhouse gases influence climate. Because this work and previous work^{7,8,18,21} have shown that hydrological impacts are also affected to a comparable extent by physiological forcing by CO_2 , and because most other greenhouse gases, such as CH_4 , N_2O and the chlorofluorocarbons, do not exert physiological forcing, radiative-forcing-based metrics give an incomplete indication of the relative effects of the different greenhouse gases on hydrological impacts. For example, CH_4 has a 100-year GWP of 23, suggesting that it is 23 times 'more potent' than CO_2 in influencing climate. However, CH_4 does not exert a physiological forcing (except indirectly by producing CO_2), so its effect on hydrological impacts relative to CO_2 may be considerably less than that implied by the GWP. O_3 is another example of a greenhouse gas that is likely to exert a physiological forcing, because it affects plant functioning directly as a poison. Moreover, atmospheric aerosols can affect surface evaporation through changes in the surface radiation budget²⁹, and can also affect P through changes in cloud droplet size²⁸. Other chemical species such as ammonia can affect plant physiology (and hence hydrology) but do not exert radiative forcings. Water resources and flood risk are among the most frequently cited issues of concern over climate change^{10,30}, so we consider that the conventional GWP-based concept of 'CO₂ equivalent' is incomplete and additional metrics for comparing greenhouse gases in terms of hydrological impacts are required.

METHODS SUMMARY

The perturbed-physics ensemble technique involves performing a large number of simulations with multiple climate model versions, each with different values assigned to key parameters involved in the simulation of climate processes. Here we used an ensemble of 224 different versions of the HadSM3 climate model^{12,13}, each with its own unique combination of parameter values or settings. One parameter varied within the ensemble determines whether plant physiological processes respond to atmospheric CO_2 changes (hereafter the physiological forcing switch, PF). Arbitrarily, PF was 'on' in 81 members of the ensemble (RADPHYS), and 'off' in the remaining 143 (RAD). Two simulations were performed with each ensemble member, one simulation at climatic equilibrium with pre-industrial CO_2 and the other at climatic equilibrium with doubled CO_2 . Each pair of simulations was termed one 'experiment'. In RAD, with PF 'off', plant physiological processes were simulated with the pre-industrial CO_2 concentration.

In Fig. 1d, each point shows the 20-year equilibrium mean runoff change versus the 20-year equilibrium mean precipitation change due to doubling CO_2 for one experiment. In Fig. 1a–c, the grey and black bars show percentages of experiments in the RADPHYS and RAD sub-ensembles respectively for which the runoff, precipitation or temperature changes lie between the increments shown on the x axis. Comparison between the grey and black bars lying between the same increments shows the relative frequency of a particular change in these quantities in experiments with and without PF. Further simulations RADPHYS_DV and RAD_DV were transient simulations additionally including changes in vegetation type and leaf area due to CO_2 and climate changes.

Our results rely on the accuracy of the parameterized transpiration sensitivity to CO_2 concentrations. We did not explicitly vary plant physiological parameters in our ensembles (aside from turning PF 'on' and 'off'), so we did not explore uncertainties in the response to physiological forcing to the same extent as those in the response to radiative forcing. However, our simulated reductions in transpiration under doubled CO_2 are consistent with experimental work using 'free-air CO_2 enrichment' (FACE) techniques^{5,6}. Our model provides a near-optimal fit to observed increases in continental runoff⁸ and is intermediate in the range of other models' responses of runoff to increasing CO_2 (refs 18 and 21).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 10 January 2007; accepted 21 June 2007.

1. Sellers, P. J. *et al.* Comparison of radiative and physiological effects of doubled atmospheric CO_2 on climate. *Science* **271**, 1402–1406 (1996).
2. Field, C., Jackson, R. & Mooney, H. Stomatal responses to increased CO_2 : implications from the plant to the global scale. *Plant Cell Environ.* **18**, 1214–1255 (1995).
3. Betts, R. A., Cox, P. M., Lee, S. E. & Woodward, F. I. Contrasting physiological and structural vegetation feedbacks in climate change simulations. *Nature* **387**, 796–799 (1997).
4. Cox, P. M. *et al.* The impact of new land surface physics on the GCM simulation of climate and climate sensitivity. *Clim. Dyn.* **15**, 183–203 (1999).
5. Hungate, B. A. *et al.* Evapotranspiration and soil water content in a scrub-oak woodland under carbon dioxide enrichment. *Glob. Change Biol.* **8**, 289–298 (2002).
6. Long, S. P., Ainsworth, E. A., Leakey, A. D. B., Nösberger, J. & Ort, D. R. Food for thought: lower-than-expected crop yield stimulation with rising CO_2 concentrations. *Science* **312**, 1918–1921 (2006).
7. Wigley, T. M. L. & Jones, P. D. Influences of precipitation changes and direct CO_2 effects on streamflow. *Nature* **314**, 149–152 (1985).
8. Gedney, N. *et al.* Detection of a direct carbon dioxide effect in continental river runoff records. *Nature* **439**, 835–838 (2006).
9. Arnell, N. *et al.* Hydrology and water resources. In *Climate Change 2001: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change* Ch.4, 191–234 (Cambridge Univ. Press, Cambridge, UK, 2001).
10. Warren, R. Impacts of global climate change at different annual mean global temperature increases. In *Avoiding Dangerous Climate Change* (eds Schellnhuber, H. J., Cramer, W., Nakicenovic, N., Wigley, T. & Yohe, G.) Ch.11, 93–100 (Cambridge Univ. Press, Cambridge, UK, 2006).
11. De Wit, M. & Stankiewicz, J. Changes in surface water supply across Africa with predicted climate change. *Science* **311**, 1917–1921 (2006).
12. Murphy, J. M. *et al.* Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature* **430**, 768–772 (2004).
13. Webb, M. J. *et al.* On the contribution of local feedback mechanisms to the range of climate sensitivity in two GCM ensembles. *Clim. Dyn.* **27** (1), 17–38 (2006).
14. Collatz, G. J., Ball, J. T., Grivet, C. & Berry, J. A. Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: A model that includes a laminar boundary layer. *Agric. Forest Meteorol.* **54**, 107–136 (1991).
15. Collatz, G. J., Ribas-Carbo, M. & Berry, J. A. A coupled photosynthesis-stomatal conductance model for leaves of C4 plants. *Aust. J. Plant Physiol.* **19**, 519–538 (1992).
16. Jacobs, C. *Direct Impacts of Atmospheric CO₂ Enrichment on Regional Transpiration*. PhD thesis, Wageningen Agricultural Univ. (1994).
17. Sellers, P. J., Berry, J., Collatz, G., Field, C. & Hall, F. Canopy reflectance, photosynthesis and transpiration. III. A reanalysis using enzyme kinetics–electron transport models of leaf physiology. *Remote Sensing Environ.* **42**, 187–216 (1992).
18. Cramer, W. *et al.* Global response of terrestrial ecosystem structure and function to CO_2 and climate change: results from six dynamic global vegetation models. *Glob. Change Biol.* **7**, 357–374 (2001).
19. Betts, R. A. *et al.* The role of ecosystem-atmosphere interactions in simulated Amazonian precipitation decrease and forest dieback under global climate warming. *Theor. Appl. Climatol.* **78**, 157–175 (2004).
20. Betts, R. A. Self-beneficial effects of vegetation on climate in an ocean-atmosphere General Circulation Model. *Geophys. Res. Lett.* **26**, 1457–1460 (1999).
21. Leipprand, A. & Gerten, D. Global effects of doubled atmospheric CO_2 content on evapotranspiration, soil moisture and runoff under potential natural vegetation. *Hydrol. Sci.* **51**, 171–185 (2006).

22. Betts, R. A., Cox, P. M. & Woodward, F. I. Simulated responses of potential vegetation to doubled-CO₂ climate change and feedbacks on near-surface temperature. *Glob. Ecol. Biogeogr.* **9**, 171–180 (2000).
23. Douville, H. *et al.* Importance of vegetation feedbacks in doubled-CO₂ climate experiments. *J. Geophys. Res.* **105** (D11), 14841–14861 (2000).
24. Cox, P. M. *Description of the Triffid Dynamic Global Vegetation Model*. Technical Note 24 (Met Office Hadley Centre, Bracknell, 2001); (<http://www.metoffice.gov.uk/research/hadleycentre/pubs/HCTN/index.html>).
25. Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A. & Totterdell, I. J. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature* **408**, 184–187 (2000).
26. IPCC. *Climate Change 1994: Radiative Forcing of Climate Change and an Evaluation of the IPCC IS92 Emission Scenarios* (Cambridge Univ. Press, Cambridge, UK, 1994).
27. United Nations Framework Convention on Climate Change. *UNFCCC Article 2*. Report no. UNEP/IUC/99/2 (Information Unit for Conventions, UNEP, Geneva, 1999); (<http://www.unfccc.int/resource/convkp.html>).
28. Forster, P. *et al.* in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Solomon, S. *et al.*) Ch. 2 (Cambridge Univ. Press, Cambridge, UK/New York, USA, 2007).
29. Roderick, M. J. & Farquhar, G. D. The cause of decreased pan evaporation over the past 50 years. *Science* **298**, 1410–1411 (2002).
30. IPCC. *Climate Change 2007: Impacts, Adaptation and Vulnerability. Working Group II Contribution to the Intergovernmental Panel on Climate Change, Fourth Assessment Report* (Cambridge Univ. Press, Cambridge, UK, 2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Dupre, D. Matthews, A. Nobre, C. Rye, M. Sanderson, S. Sitch and T. Wheeler for comments. This work was supported by the UK Ministry of Defence project “Defence and Security Implications of Climate Change” and the Climate Prediction Programme of the UK Department for Environment, Food and Rural Affairs. P.M.C. and C.H. were supported by the UK Natural Environment Research Council.

Author Contributions R.A.B. proposed the study, performed the dynamic vegetation simulations and led the analysis and writing. D.L.H. performed statistical analysis of the ensemble simulations and contributed expertise on field experiments on plant physiology. P.D.F. analysed the dynamic vegetation simulations. P.M.C. developed the MOSES and TRIFFID models and contributed to the interpretation. C.D.J., N.G., C.H. and O.B. contributed to the analysis and provided further expertise on modelling plant physiology, hydrology and land–atmosphere interactions. M.C., D.M.H.S. and M.J.W. designed and performed the ensemble simulations and advised on their interpretation. All co-authors contributed to the text.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.A.B. (richard.betts@metoffice.gov.uk).

METHODS

The perturbed-physics ensemble technique has previously been used to begin to quantify the uncertainties in climate predictions arising from uncertainties in the values assigned to key parameters input to the HadSM3 climate model¹². Most of the parameters are numeric quantities which have a range of possible values but which are assigned a particular value in the standard version of HadSM3^{12,13}. Some parameters act as logical switches which determine whether particular processes are represented within the model; a particular example is the switch for plant physiological responses to changes in atmospheric CO₂ (PF), which is enabled in the standard version of HadSM3 but can be disabled if required. If this process is disabled, the plant physiological process calculations are assigned a fixed CO₂ concentration as an input term, so any changes in plant physiology are a response to changes in meteorological and hydrological changes only.

Each of the 224 slightly different versions of HadSM3 had its own unique combination of parameter values or settings. These combinations can be thought of as points in a 31-dimensional parameter space defined by the values of the numeric parameters and the 'on/off' settings of the logical switches. The values of numeric parameters were varied between maximum and minimum plausible values as judged by experts^{12,13}, with intermediate values as used in the standard version of HadSM3 also being used. The combinations of parameter values were selected partly to provide representative coverage of this parameter space, and not to examine the responses to any particular parameter. For the purposes of generating the ensemble, the two settings of the physiological forcing switch are considered equally likely. The ensemble was not designed to systematically examine the effect of switching PF "on" or "off", so there were no pairs of ensemble members which differed only in the setting of the PF switch individual. That is, the members of RADPHYS had no directly parallel equivalents in RAD.

While the two sub-ensembles therefore do not provide a perfect controlled experiment for examining the effects of physiological forcing, because combinations of other parameter values also differ, there is sufficient evidence to suggest that these other differences exert random effects on the climate and do not introduce any systematic bias in runoff or its response to doubling CO₂. The numbers of members in the RADPHYS and RAD sub-ensembles were sufficient to ensure that the differences between the two sub-ensemble mean changes were statistically significant at 1% for runoff and temperature and 5% for precipitation, but there was no statistically significant bias in the control simulations. Moreover, previous work⁴ with a single pair of simulations with the same atmosphere–land model provides a controlled experiment supporting the results of our ensemble. That work⁴ showed an increase in runoff of 26 ± 7 and $11 \pm 7 \text{ kg m}^{-2} \text{ yr}^{-1}$ with and without physiological forcing respectively (with standard deviations in that study referring to annual means in a single simulation).

An intercomparison of the responses of six vegetation models including our own¹⁹ showed that all the models produced increased runoff due to physiological forcing by CO₂, and our model was intermediate in the range of responses. One recent FACE experiment⁶ showed a localized warming of 1.4 K over a 20-m-diameter plot as a result of reduced transpiration due to increasing CO₂ from 380 p.p.m. to 550 p.p.m. The mean difference in global land average warming between RADPHYS and RAD was 0.96 K, for an increase in CO₂ from 280 to 540 p.p.m. This smaller increase in warming for a larger CO₂ rise may partly reflect the fact that our global land average results include areas with no vegetation.

LETTERS

Dating the origin of the Orchidaceae from a fossil orchid with its pollinator

Santiago R. Ramírez¹, Barbara Gravendeel², Rodrigo B. Singer³, Charles R. Marshall^{1,4} & Naomi E. Pierce¹

Since the time of Darwin¹, evolutionary biologists have been fascinated by the spectacular adaptations to insect pollination exhibited by orchids. However, despite being the most diverse plant family on Earth², the Orchidaceae lack a definitive fossil record and thus many aspects of their evolutionary history remain obscure. Here we report an exquisitely preserved orchid pollinarium (of *Meliorchis caribea* gen. et sp. nov.) attached to the mesoscutellum of an extinct stingless bee, *Proplebeia dominicana*, recovered from Miocene amber in the Dominican Republic, that is 15–20 million years (Myr) old³. This discovery constitutes both the first unambiguous fossil of Orchidaceae⁴ and an unprecedented direct fossil observation of a plant–pollinator interaction^{5,6}. By applying cladistic methods to a morphological character matrix, we resolve the phylogenetic position of *M. caribea* within the extant subtribe Goodyerinae (subfamily Orchidoideae). We use the ages of other fossil monocots and *M. caribea* to calibrate a molecular phylogenetic tree of the Orchidaceae. Our results indicate that the most recent common ancestor of extant orchids lived in the Late Cretaceous (76–84 Myr ago), and also suggest that the dramatic radiation of orchids began shortly after the mass extinctions at the K/T boundary. These results further support the hypothesis of an ancient origin for Orchidaceae.

Family Orchidaceae Juss., 1789
Subtribe Goodyerinae Klotzsch, 1846
Meliorchis caribea gen. et sp. nov.

Etymology. The generic name alludes to the plant's pollination mode by meliponine bees and incorporates the Greek name of an orchid (orchis: testicle). The specific epithet *caribea* refers to the Caribbean region.

Holotype. Museum of Comparative Zoology (Harvard University), catalogue number MCZ-31141.

Horizon and locality. Specimen was excavated in the year 2000 from a mine located east of Santiago, Cordillera Septentrional, Dominican Republic. Lignite and sandy clay beds, Early to Middle Miocene (15–20 Myr old; ref. 3).

Diagnosis. The species is separated from other members of Goodyerinae by the bent anther, large angular massulae (~100 per pollinarium), and tightly packed pollen units (20 × 20 µm). The amber piece (20 × 14 × 5 mm) contains a single inclusion of *Meliorchis caribea*. Two complete pollinia (each ~1,000 × 500 µm), belonging to a single pollinarium, are firmly attached to the mesoscutellum of a worker bee, *Proplebeia dominicana*⁷ (Fig. 1a). The tapering pollinia consist of >100 loosely packed angular massulae (~200 × 100 µm, Fig. 1b), each of which encapsulates several tetrads; obovoid pollen units are tightly packed.

These pollinarium features are found only in the Orchidoideae⁸. A survey of herbarium specimens of all Neotropical genera within this

subfamily showed that the size, shape and ornamentation of the fossil closely resemble those of modern members of the subtribe Goodyerinae, particularly the genera *Kreodanthus* and *Microchilus* (Supplementary Table 1). In addition, the position of the pollinarium on the fossilized bee enables us to make inferences about unique aspects of the flowers of *Meliorchis*, even in the absence of fossil flowers. Whereas in living Goodyerinae the pollinarium normally is attached to the mouthparts of pollinating bees⁹ (Fig. 2a), the

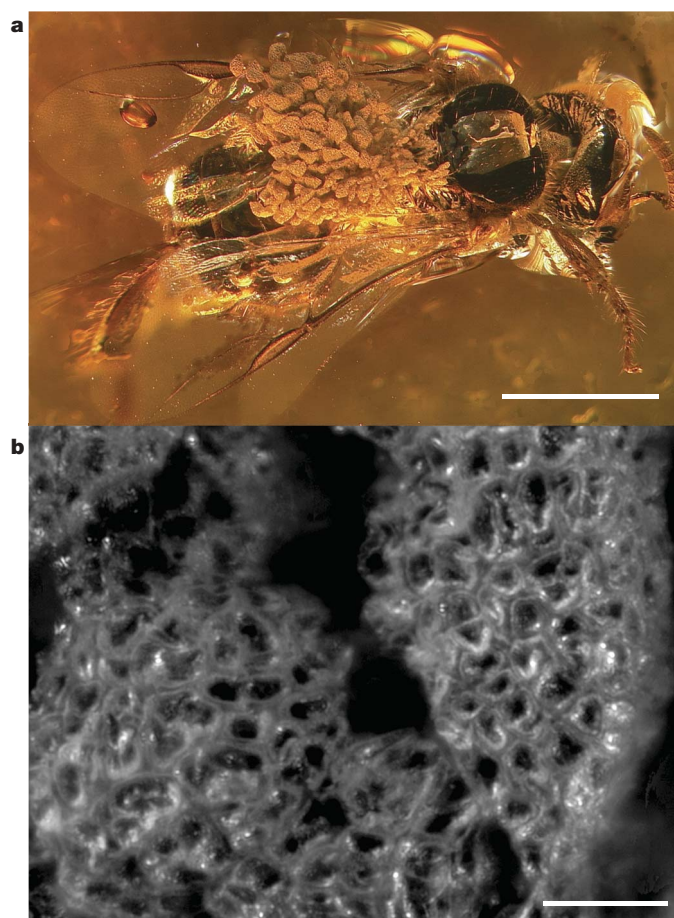


Figure 1 | Holotype of *Meliorchis caribea* gen. et sp. nov. This orchid pollinarium, carried by a worker stingless bee (*Proplebeia dominicana*), is preserved in amber from the Dominican Republic and represents the first definitive fossil record for the family Orchidaceae. **a**, General view of encapsulated specimen (scale bar, 1,000 µm). **b**, Detailed view of the pollinia surface showing pollen units (scale bar, 50 µm).

¹Museum of Comparative Zoology, Harvard University, 26 Oxford St., Cambridge, Massachusetts 02138, USA. ²Nationaal Herbarium Nederland, Universiteit Leiden, P.O. Box 9514, Leiden, The Netherlands. ³Depto Botânica, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves 9500, RS 91501-970, Porto Alegre, Brasil.

⁴Department of Earth and Planetary Sciences, Harvard University, 20 Oxford St., Cambridge, Massachusetts 02138, USA.

pollinarium of *Meliorchis* is attached to the mesoscutellum (dorsal surface of the thorax) of worker bees of *P. dominicana*. This indicates that the flower of *M. caribea* was gullet-shaped, and, rather than the bee probing the lip of the flower with its tongue as in modern Goodyerinae (Fig. 2a), the anterior part of the bee would have had to enter the flower completely (Fig. 2b).

Because evidence of plant–pollinator interactions is exceedingly rare in the fossil record, our current knowledge of ancient pollination is indirectly inferred from specialized morphological features of fossilized insects^{10–12} and flowers^{13–15}. In addition, records of pollen grains on fossil insects and in coprolites provide circumstantial evidence for ancient insect–flower interactions^{5,6,10,12,14}, although these observations—with the exception of amber-preserved fig wasps carrying fig pollen⁶—do not exclude the possibility of flower visitation without pollination⁵. In contrast, because in most orchids the staminal filaments are fused to the style, the anatomical match required for a pollinator to remove the pollinarium is nearly identical to that necessary for its subsequent delivery (Fig. 2). Thus, *P. dominicana* bee workers were almost certainly pollinators of flowers of *M. caribea*. Because modern stingless bees pollinate numerous rainforest angiosperms¹⁶, including several tropical orchid species¹⁷, this fossil shows that adaptation by tropical orchids to specialized pollinators occurred at least as far back as the Miocene.

To explore the phylogenetic position of *Meliorchis* in relation to Modern orchid taxa, we constructed a morphological character matrix consisting of 25 characters and 15 taxa adapted from a previous study¹⁸ (see Supplementary Methods for details). Heuristic tree searches optimized by maximum parsimony yielded 129 equally short trees, all of which supported monophyly of both the subfamily Orchidoideae and the subtribe Goodyerinae (Fig. 3). The position of *Meliorchis* within Goodyerinae is supported by a bootstrap of 91%. Of the 129 recovered trees, none supported *Meliorchis* as a sister clade to the rest of the Goodyerinae genera. Together, these results indicate that *Meliorchis* represents a differentiated lineage within extant Goodyerinae. On the basis of estimated ages of Dominican amber³, a minimum age of 15–20 Myr can be assigned to the subtribe Goodyerinae.

Previously published putative orchid fossils have lacked diagnostic characters that would definitively assign them to Orchidaceae^{4,19}. In fact, in a thorough review of all known specimens, it was concluded that Orchidaceae have ‘no positive or useful fossil record’⁴. This absence in the fossil record, most likely owing to their non-diagnostic leaves and lack of wind-dispersed pollen, has spurred considerable disagreement regarding orchids’ age of origin and timing of

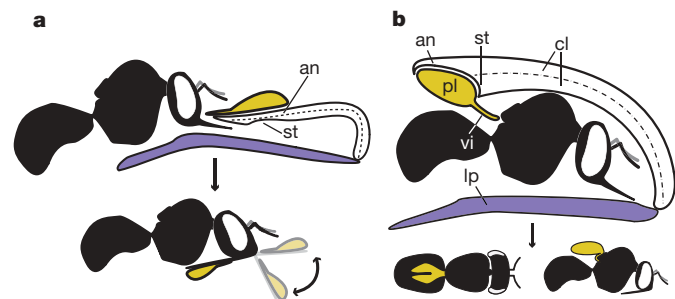


Figure 2 | Morphology and pollinarium placement of modern Goodyerinae and hypothetical reconstruction of floral morphology of *Meliorchis caribea*.

a, The parallel lip (lp) and column (cl) and the erect anther (an) of extant Goodyerinae typically result in the pollinarium (pl) attachment on the pollinator's mouthparts. **b**, The attachment of the pollinarium to the mesoscutellum (dorsal surface of thorax) of a worker bee is only possible when the lip and column of the flower are parallel but the anther is bent. Under this scenario, the distance between the lip and the column must be ~2.5 mm to enable a *P. dominicana* worker to crawl into the flower and remove the pollinarium with its mesoscutellum as it retreats; st, stigma; vi, viscidium.

diversification. Whereas orchids' highly specialized pollination mechanisms, epiphytism and absence in fossil deposits were cited by early workers in support of a recent age^{4,20,21}, their worldwide distribution² and basal placement in the order Asparagales²² suggest an older age. Indeed, three recent molecular clock studies that broadly sampled angiosperm clades (including a few orchid representatives) obtained radically different age estimates for the Orchidaceae, ranging from ~26 Myr old²³ and ~40 Myr old²⁴ to ~110 Myr old²⁵. Such age discrepancies are most likely due to under-represented sampling and absence of internal calibration points. We here use both the age and phylogenetic position of *M. caribea* and other fossil monocots to estimate the timing of diversification for Orchidaceae.

We calibrated a molecular phylogenetic tree of Orchidaceae by implementing a relaxed-clock model through penalized likelihood and non-parametric rate smoothing (NPRS). We built a molecular phylogenetic tree of Orchidaceae that was based on plastid DNA sequences obtained from GenBank for 55 orchid genera representing all major lineages in the family, and five basal Asparagales genera as outgroup taxa. Our divergence time estimates using penalized likelihood suggest that extant Orchidaceae shared a most recent common ancestor in the Late Cretaceous, 76 ± 5 to 84 ± 6 Myr ago, depending on whether we use the oldest or youngest estimates of the ages of the fossils used to calibrate the relaxed molecular clock (Fig. 4). Similarly, age estimates obtained using NPRS suggest that crown Orchidaceae shared a common ancestor 76 ± 4 to 83 ± 4 Myr ago. Our results also suggest that stem lineages of all five orchid subfamilies were present early in the evolutionary history of Orchidaceae, before the end of the Cretaceous, ~65 Myr ago (Fig. 4). The extant lineages of the two largest orchid subfamily clades (Orchidoideae and Epidendroideae), which together encompass >95% of the living orchid species, began to diversify early in the Tertiary, although more thorough taxonomic sampling could result in older age estimates of their common ancestor.

The discovery of *Meliorchis caribea* and the internally calibrated molecular clock analyses presented here reject the hypothesis of a relatively recent (Eocene or younger) origin of Orchidaceae^{4,21}.

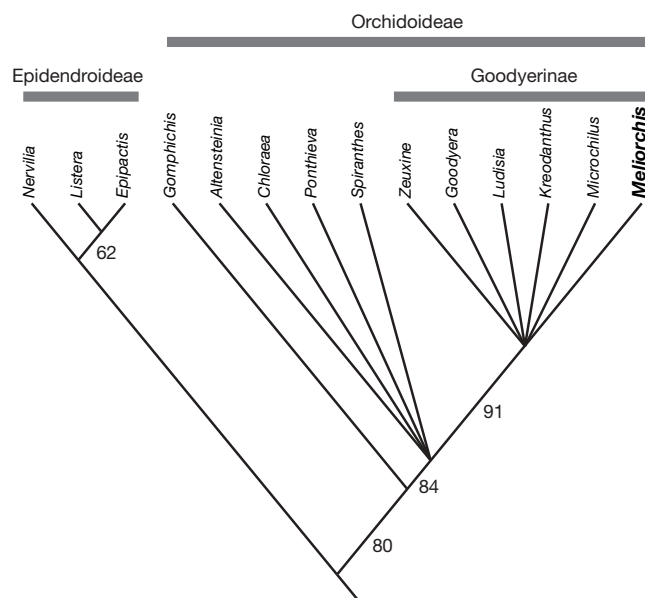


Figure 3 | Cladogram showing the estimated position of *Meliorchis* among modern clades in the orchid subfamily Orchidoideae. A strict consensus of the 129 shortest trees (tree length = 42, consistency index = 0.619, retention index = 0.660) obtained using 25 morphological characters for 15 taxa; values beside nodes correspond to bootstrap percentages (1,000 replicates). None of the shortest trees recovered *Meliorchis* as sister to all the other Goodyerinae included.

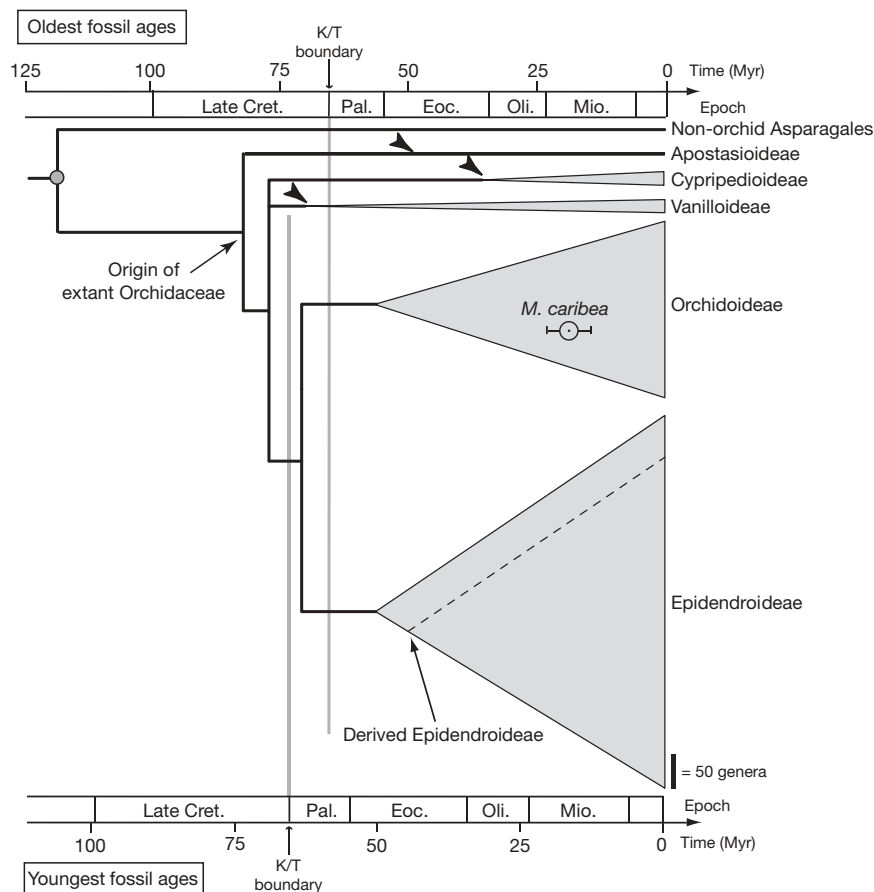


Figure 4 | Fossil-calibrated molecular clock chronogram of the family Orchidaceae, based on ~3 kilobases of plastid DNA (*matK* and *rbcL*). The relative size of each clade is proportional to the number of genera described in each orchid subfamily. Crown ages of small clades are indicated with arrow heads. Two sets of dates were used to calculate orchid divergence times: the oldest and youngest estimates of the ages of the fossils. The age boundaries of *M. caribea* (15–20 Myr old) relative to each timescale

Instead, our results favour the hypothesis of an ancient (Late Cretaceous)^{22,25} origin of extant Orchidaceae, but at the same time support a Tertiary radiation of the most diverse epiphytic clades. Our age estimates are younger than the oldest proposed for the family by previous studies²⁵, but we note that our age calculations should be regarded as minimum estimates, which could be pushed back with additional fossil discoveries. Our scenario corresponds to that previously proposed^{2,22}, is consistent with the observed disjunct pantropical distributions of the subfamily clades and the early-splitting genera (for example, *Vanilla*), and reinforces the possibility of a Late Cretaceous biotic exchange between tropical continents.

METHODS

Colour photomicrographs were taken with a JVC digital camera (KYF75U) mounted on a MZ16 Leica dissecting scope; black and white micrographs were taken with a Retiga EXi digital camera mounted on a Leica Leitz-dmrb compound microscope (objective $\times 40$). In both cases, 10 sequential shots at different focal depths were processed with the Auto-Montage software (Synchroscopy, 2002) to produce a single composite image.

The phylogenetic position of *Meliorchis* was explored using morphological characters from flowers, pollinaria and pollen micro-morphology, all of which were directly observable or inferable from the type specimen of *M. caribea*. We treated all character states as unordered and weighted them equally. Because *Meliorchis* unambiguously belongs to the subfamily Orchidoideae, we only included representative genera from this group. We selected outgroup taxa on the basis of previous studies that used both morphological¹⁸ and molecular²⁶ data. Heuristic tree searches were performed via maximum parsimony with

correspond to the distance between the circle's centre and the vertical bar. Additional monocot fossil records outside Orchidaceae were used to calibrate the root of the tree (node indicated by filled circle). Branch lengths were optimized under the maximum likelihood model of sequence evolution GTR+ Γ +I using a 95% majority-rule consensus tree (see Supplementary Information); node ages were estimated using a penalized likelihood method²⁹.

the TBR algorithm (100 random addition replicates). A total of 1,000 replicates were run to estimate bootstrap support; all analyses were performed in PAUP* v.4.0b.

Consensus phylogenetic trees of bayesian analyses were obtained with the software MrBayes v3.1.1 (for details, see Supplementary Materials). Our topologies agree with those obtained by previous studies^{27,28}. Divergence times were calculated by penalized likelihood and NPRS, using the truncated Newton algorithm in the software r8s v 1.71²⁹. Two sets of dates were used, corresponding to the youngest and oldest estimates of the ages of the fossils used as node age constraints. We applied (1) the age of *Meliorchis* (15–20 Myr old; ref. 3) as a minimum age for the monophyletic Goodyerinae; (2) the age of the oldest known Asparagales (93–105 Myr old, see Supplementary Methods for details) as a minimum age constraint at the root of the tree; and (3) the age of the oldest known fossil monocot as the maximum age at the root of the tree (110–120 Myr old; ref. 30).

Received 17 January; accepted 21 June 2007.

1. Darwin, C. *On the Various Contrivances by which British and Foreign Orchids are Fertilised by Insects, and on the Good Effects of Intercrossing* (J. Murray, London, 1862).
2. Dressler, R. L. *The Orchids: Natural History and Classification* (Harvard Univ. Press, Cambridge, Massachusetts, 1981).
3. Iturralde-Vinent, M. E. & MacPhee, R. D. E. Age and paleogeography of Dominican amber. *Science* **273**, 1850–1852 (1996).
4. Schmid, R. & Schmid, M. J. in *Orchid Biology: Reviews and Perspectives* Vol. 1 (ed. Arditti J.) 17–45 (Cornell Univ. Press, London, 1977).
5. Grimaldi, D. & Engel, M. S. *Evolution of the Insects* (Cambridge Univ. Press, New York, 2005).
6. Peñalver, E., Engel, M. S. & Grimaldi, D. Fig wasps in Dominican amber (Hymenoptera: Agaonidae). *Am. Mus. Novit.* **3541**, 1–16 (2006).

7. Camargo, J. M. F., Grimaldi, D. & Pedro, S. R. M. The extinct fauna of stingless bees (Hymenoptera: Apidae: Meliponini) in Dominican amber: Two new species and redescription of the male of *Proplebeia dominicana* (Wille and Chandler). *Am. Mus. Novit.* **3293**, 1–24 (2000).
8. Freudenstein, J. V. & Rasmussen, F. N. Sectile pollinia and relationships in Orchidaceae. *Plant Syst. Evol.* **205**, 125–146 (1997).
9. Singer, R. B. & Sazima, M. Flower morphology and pollination mechanism in three sympatric Goodyerinae orchids from southeastern Brazil. *Ann. Bot. (Lond.)* **88**, 989–997 (2001).
10. Poinar, G. O. & Danforth, B. N. A fossil bee from Early Cretaceous Burmese amber. *Science* **314**, 614 (2006).
11. Ren, D. Flower-associated Brachycera flies as fossil evidence for Jurassic angiosperm origins. *Science* **280**, 85–88 (1998).
12. Grimaldi, D. The co-radiations of pollinating insects and angiosperms in the Cretaceous. *Ann. Mo. Bot. Gard.* **86**, 373–406 (1999).
13. Crepet, W. L., Friis, E. M., Nixon, K. C., Lack, A. J. & Jarzembowski, E. A. Fossil evidence for the evolution of biotic pollination. *Phil. Trans. R. Soc. London. B* **333**, 187–195 (1991).
14. Crepet, W. L. Some aspects of the pollination biology of Middle Eocene angiosperms. *Rev. Palaeobot. Palynol.* **27**, 213–238 (1979).
15. Gandolfo, M. A., Nixon, K. C. & Crepet, W. L. Cretaceous flowers of Nymphaeaceae and implications for complex insect entrapment pollination mechanisms in early Angiosperms. *Proc. Natl Acad. Sci. USA* **101**, 8056–8060 (2004).
16. Heard, T. A. The role of stingless bees in crop pollination. *Annu. Rev. Entomol.* **44**, 183–206 (1999).
17. Roubik, D. W. Deceptive orchids with Meliponini as pollinators. *Plant Syst. Evol.* **222**, 271–279 (2000).
18. Freudenstein, J. V. & Rasmussen, F. N. What does morphology tell us about orchid relationships?—A cladistic analysis. *Am. J. Bot.* **86**, 225–248 (1999).
19. Herendeen, P. S. & Crane, P. S. in *Monocotyledons: Systematics and Evolution* (eds Rudall, P. J., Cribb, P. J., Cutler, D. F. & Humphries, C. J.) 1–21 (Royal Botanic Gardens, Kew, 1995).
20. Crepet, W. L. Insect pollination: a paleontological perspective. *Bioscience* **29**, 102–107 (1979).
21. Labandeira, C. C. Paleobiology: how old is the flower and the fly? *Science* **280**, 57–59 10.1126/science.280.5360.57 (1998).
22. Chase, M. W. in *Genera Orchidacearum* Vol. 2 (eds Pridgeon, A. M., Cribb, P. J., Chase, M. W. & Rasmussen, F. N.) 1–5 (Oxford Univ. Press, New York, 2001).
23. Wikström, N., Savolainen, V. & Chase, M. W. Evolution of the angiosperms: calibrating the family tree. *Proc. R. Soc. Lond. B* **268**, 2211–2220 (2001).
24. Bremer, K. Early Cretaceous lineages of monocot flowering plants. *Proc. Natl Acad. Sci. USA* **97**, 4707–4711 (2000).
25. Janssen, T. & Bremer, K. The age of major monocot groups inferred from 800+ *rbcl* sequences. *Bot. J. Linn. Soc.* **146**, 385–398 (2004).
26. van der Berg, C. et al. An overview of the phylogenetic relationships within Epidendroideae inferred from multiple DNA regions and recircumscription of Epidendreae and Arethuseae (Orchidaeeae). *Am. J. Bot.* **92**, 613–624 (2005).
27. Cameron, K. M. et al. A phylogenetic analysis of the Orchidaceae: evidence from *rbcl* nucleotide sequences. *Am. J. Bot.* **86**, 208–224 (1999).
28. Freudenstein, J. V. et al. An expanded plastid DNA phylogeny of Orchidaceae and analysis of jackknife branch support strategy. *Am. J. Bot.* **91**, 149–157 (2004).
29. Sanderson, M. J. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301–302 (2003).
30. Friis, E. M., Pedersen, K. R. & Crane, P. R. Araceae from the Early Cretaceous of Portugal: Evidence on the emergence of monocotyledons. *Proc. Natl Acad. Sci. USA* **101**, 16565–16570 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Y. Goldman for facilitating access to the amber inclusion discussed here, G. Romero for his assistance in the examination of herbarium specimens, and G. Alpert and D. Smith for assistance in the production of the fossil microphotographs. We thank B. Archibald, C. Bell, M. Chase, A. Knoll, R. van der Ham, D. Hewitt, C. Jaramillo, M. Patten, E. Pringle, J. Pringle and T. Quental for useful comments. This research was sponsored by grants from the Barbour Fund (Museum of Comparative Zoology) and the National Science Foundation (DDIG) to S.R.R. and N.E.P., and a grant from the Fulbright Junior Scholar programme to B.G.

Author Contributions S.R.R., B.G. and N.E.P. procured and curated the specimen. S.R.R., B.G. and R.B.S. reviewed herbaria specimens, analysed ancestral floral morphology, and coded and analysed morphological characters. S.R.R. and C.R.M. designed dating approaches and considered their interpretation. S.R.R. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.R.R. (sramirez@oeb.harvard.edu).

LETTERS

Mobility promotes and jeopardizes biodiversity in rock–paper–scissors games

Tobias Reichenbach¹, Mauro Mobilia¹ & Erwin Frey¹

Biodiversity is essential to the viability of ecological systems. Species diversity in ecosystems is promoted by cyclic, non-hierarchical interactions among competing populations. Central features of such non-transitive relations are represented by the ‘rock–paper–scissors’ game, in which rock crushes scissors, scissors cut paper, and paper wraps rock. In combination with spatial dispersal of static populations, this type of competition results in the stable coexistence of all species and the long-term maintenance of biodiversity^{1–5}. However, population mobility is a central feature of real ecosystems: animals migrate, bacteria run and tumble. Here, we observe a critical influence of mobility on species diversity. When mobility exceeds a certain value, biodiversity is jeopardized and lost. In contrast, below this critical threshold all subpopulations coexist and an entanglement of travelling spiral waves forms in the course of time. We establish that this phenomenon is robust; it does not depend on the details of cyclic competition or spatial environment. These findings have important implications for maintenance and temporal development of ecological systems and are relevant for the formation and propagation of patterns in microbial populations or excitable media.

The remarkable biodiversity present in ecosystems confounds a naive interpretation of darwinian evolution in which interacting species compete for limited resources until only the fitter species survives. As a striking example, consider that a 30 g sample of soil from a Norwegian forest is estimated to contain some 20,000 common bacterial species⁶. Evolutionary game theory^{7–9}, in which the success of one species relies on the behaviour of others, provides a useful framework in which to investigate co-development of populations theoretically. In this context, the rock–paper–scissors game has emerged as a paradigm to describe species diversity^{1–5,10–12}. If three subpopulations interact in this non-hierarchical way, we intuitively expect that diversity may be preserved: Each species dominates another only to be outperformed by the remaining one in an endlessly spinning wheel of species chasing species.

Communities of subpopulations exhibiting such dynamics have been identified in numerous ecosystems, ranging from coral reef invertebrates¹³ to lizards in the inner Coast Range of California¹⁴. In particular, recent experimental studies using microbial laboratory cultures have been devoted to the influence of spatial structure on time development and coexistence of species^{3,15}. Investigating three strains of colicinogenic *Escherichia coli* in different environments, it has been shown that cyclic dominance alone is not sufficient to preserve biodiversity. Only when the interactions between individuals are local (for example, bacteria arranged on a Petri dish) can spatially separated domains dominated by one subpopulation form and lead to stable coexistence^{1,3}.

Here we show that biodiversity is affected drastically by spatial migration of individuals, a ubiquitous feature of real ecosystems.

Migration competes with local interactions such as reproduction and selection, thereby mediating species preservation and biodiversity. For low values of mobility, the temporal development is dominated by interactions among neighbouring individuals, resulting in the long-term maintenance of species diversity. In contrast, when species mobility is high, spatial homogeneity results and biodiversity is lost. Interestingly, a critical value of mobility sharply delineates these two scenarios. We obtain concise predictions for the fate of the ecological system as a function of species mobility, thereby gaining a comprehensive understanding of its biodiversity.

The influence of mobility on species coexistence was previously studied within the framework of coupled habitat patches (‘island models’)^{16–19}. In particular, Levin considered an idealized two-patch system and observed a critical mobility for stable coexistence¹⁶. Other models comprising many spatially arranged patches were shown to facilitate pattern formation^{17,18}. Because often in nature spatial degrees of freedom vary continuously (for example, bacteria can visit the entire area of a Petri dish), we relax the simplifying assumption of habitat patches and consider continuous spatial distribution of individuals. Moreover, as an inherent feature of real ecosystems and in contrast to previous deterministic investigations^{16–19}, we explicitly take the stochastic character of the interactions among the populations into account. Such interacting particle systems, where individuals are discrete and space is treated explicitly, have already been considered in ecological contexts^{1,2,4,5,20}. The behaviour of these models often differs from what is inferred from deterministic reaction–diffusion equations, or from interconnected patches²⁰. In the case of cyclic competition, such stochastic spatial systems have been shown to allow for stable coexistence of all species^{1,2,4,5} when individuals are static. Here we explore the novel features emerging from individuals’ mobility.

Consider mobile individuals of three subpopulations (referred to as A, B and C), arranged on a spatial lattice, where they can only interact with nearest neighbours. For the possible interactions, we consider a version of the rock–paper–scissors game, namely a stochastic spatial variant of the model introduced in 1975 by May and Leonard¹⁰ (see Methods). Schematic illustrations of the model’s dynamics are provided in Fig. 1. The basic reactions comprise selection and reproduction processes, which occur at rates σ and μ , respectively. Individuals’ mobility stems from the possibility that two neighbouring individuals will swap their position (at rate ε) and will move to an adjacent empty site: hence, individuals randomly migrate on the lattice. We define the length of the square lattice as the size unit, and denote by N the number of sites. Within this setting, and applying the theory of random walks²¹, the typical area explored by one mobile individual per unit time is proportional to $M = 2\varepsilon N^{-1}$, which we refer to as the mobility. The interplay of the latter with selection and reproduction processes sensitively

¹Arnold Sommerfeld Center for Theoretical Physics (ASC) and Center for NanoScience (CeNS), Department of Physics, Ludwig-Maximilians-Universität München, Theresienstrasse 37, D-80333 München, Germany.

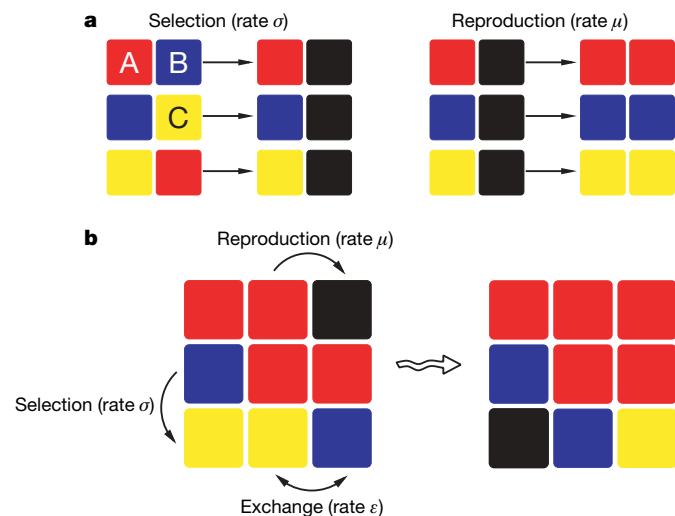


Figure 1 | The rules of the stochastic model. Individuals of three competing species A (red), B (blue), and C (yellow) occupy the sites of a lattice. **a**, They interact with their nearest neighbours through selection or reproduction, both of which reactions occur as Poisson processes at rates σ and μ , respectively. Selection reflects cyclic dominance: A can kill B, yielding an empty site (black). In the same way, B invades C, and C in turn outcompetes A. Reproduction of individuals is only allowed on empty neighbouring sites, to mimic a finite carrying capacity of the system. We also endow individuals with mobility: at exchange rate ε , they are able to swap position with a neighbouring individual or hop onto an empty neighbouring site (exchange). **b**, An example of the three processes, taking place on a 3×3 square lattice.

determines whether species can coexist on the lattice or not, as discussed below.

We performed extensive computer simulations of the stochastic system (see Methods) and typical snapshots of the steady states are reported in Fig. 2. When the mobility of the individuals is low, we find that all species coexist and self-arrange by forming patterns of moving spirals. With increasing mobility M , these structures grow in size, and disappear for large enough M . In the absence of spirals, the system adopts a uniform state where only one species is present, while the others have died out. Which species remains is subject to a random process, all species having equal chances to survive in our model.

We obtain concise predictions on the stability of three-species coexistence by adapting the concept of extensivity from statistical physics (see Supplementary Notes). We consider the typical waiting time T until extinction occurs, and its dependence on the system size N . If $T(N) \propto N$, the stability of coexistence is marginal¹². Conversely, longer (shorter) waiting times scaling with higher (lower) powers of N indicate stable (unstable) coexistence. These three scenarios can be distinguished by computing the probability P_{ext} that two species have gone extinct after a waiting time $t \propto N$. In Fig. 2, we report the dependence of P_{ext} on the mobility M . For illustration, we have considered equal reaction rates for selection and reproduction, and, without loss of generality, set the time-unit by fixing $\sigma = \mu = 1$. With increasing system size N , a sharpened transition emerges at a critical value $M_c = (4.5 \pm 0.5) \times 10^{-4}$ for the fraction of the entire lattice area explored by an individual in one time-unit. Below M_c , the extinction probability P_{ext} tends to zero as the system size increases, and coexistence is stable (implying super-persistent

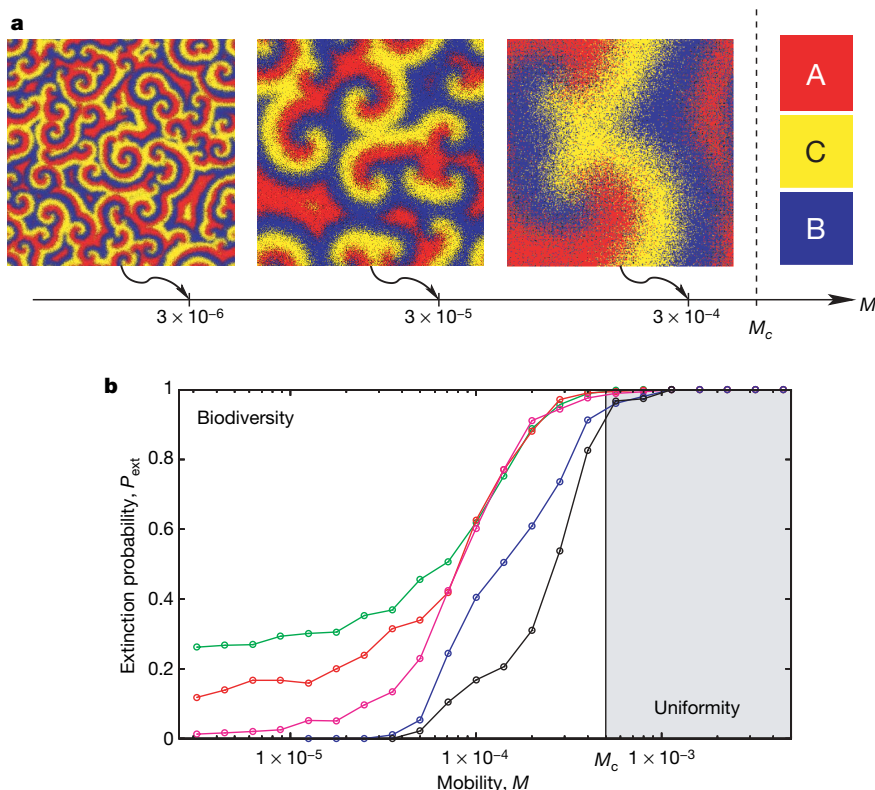


Figure 2 | The critical mobility M_c . Mobility below the value M_c induces biodiversity; while it is lost above that threshold. **a**, We show snapshots obtained from lattice simulations of typical states of the system after long temporal development (that is, at time $t \propto N$) and for different values of M (each colour represents one of the three species and black dots indicate empty spots). With increasing M (from left to right), the spiral structures grow, and outgrow the system size at the critical mobility M_c . Then coexistence of all three species is lost and uniform populations remain

(right). **b**, Quantitatively, we have considered the extinction probability P_{ext} that, starting with randomly distributed individuals on a square lattice, the system has reached an absorbing state after a waiting time $t = N$. We compute P_{ext} as a function of the mobility M (and $\sigma = \mu = 1$), and show results for different system sizes: $N = 20 \times 20$ (green), $N = 30 \times 30$ (red), $N = 40 \times 40$ (purple), $N = 100 \times 100$ (blue), and $N = 200 \times 200$ (black). As the system size increases, the transition from stable coexistence ($P_{\text{ext}} = 0$) to extinction ($P_{\text{ext}} = 1$) sharpens at a critical mobility $M_c = (4.5 \pm 0.5) \times 10^{-4}$.

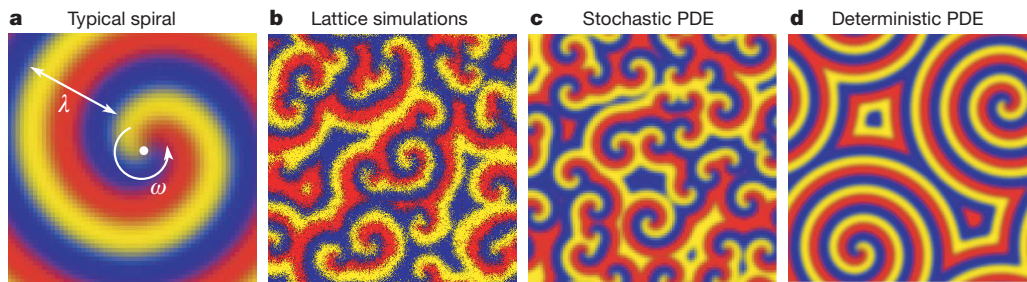


Figure 3 | Spiralling patterns. **a**, Typical spiral (schematic). It rotates around the origin (white dot) at a frequency ω and possesses a wavelength λ . **b**, In our lattice simulations, when the mobility of individuals lies below the critical value, all three species coexist, forming mosaics of entangled, rotating spirals (each colour represents one of the species and black dots indicate empty spots). **c**, We have found that the system's development can aptly be described by stochastic PDE. In the case of lattice simulations and stochastic PDE, internal noise acts as a source of local inhomogeneities and ensures the robustness of the dynamical behaviour: the spatio-temporal

transients²²; see Supplementary Notes). On the other hand, above the critical mobility, the extinction probability approaches 1 for large system size, and coexistence is unstable. One of our central results is that we have identified a mobility threshold for biodiversity:

There exists a critical value M_c such that a low mobility $M < M_c$ guarantees coexistence of all three species, while $M > M_c$ induces extinction of two of them, leaving a uniform state with only one species.

To give a biological illustration of this statement, let us consider colicinogenic strains of *E. coli* growing on a Petri dish³. In this setting, ten bacterial generations have been observed in 24 h, yielding selection and reproduction rates of about ten per day. The typical size of a Petri dish is roughly 10 cm, so we have evaluated the critical mobility to be about $5 \times 10^2 \mu\text{m}^2 \text{s}^{-1}$. Comparing that estimate to the mobility of *E. coli*, we find that it can, by swimming and tumbling in super-soft agar, explore areas of more than $10^3 \mu\text{m}^2 \text{s}^{-1}$ (ref. 23). This value can be considerably lowered by increasing the agar concentration.

When the mobility is low ($M < M_c$), the interacting subpopulations exhibit fascinating patterns, as illustrated by the snapshots of Fig. 2. The emerging reactive states are formed by an entanglement of spiral waves, characterizing the competition among the species which endlessly hunt each other, as illustrated in Supplementary Videos 1 and 2 (see also Supplementary Discussion). Formation of this type of patterns has been observed in microbial populations, such as myxobacteria aggregation²⁴ or multicellular *Dictyostelium* mounds²⁵, as well as in cell signalling and control²⁶. Remarkably, a mathematical description and techniques borrowed from the theory of stochastic processes²⁷ allow us to obtain these complex structures by means of stochastic partial differential equations (PDE), see Fig. 3 and Methods. Furthermore, recasting the dynamics in the form of a complex Ginzburg–Landau equation^{28,29} allows us to obtain analytical expressions for the spirals' wavelength λ and frequency (see Supplementary Notes). These results, up to a constant prefactor, agree with those of numerical computations, and will be published elsewhere (manuscript in preparation).

As shown in Fig. 2, the spirals' wavelength λ rises with the individuals' mobility. Our analysis reveals that the wavelength is proportional to \sqrt{M} (see Supplementary Notes). This relation holds up to the mobility M_c , where a critical wavelength λ_c is reached. For mobilities above the threshold M_c , the spirals' wavelength λ exceeds the critical value λ_c and the patterns outgrow the system size, causing the loss of biodiversity (see Fig. 2). We have found λ_c to be universal, that is, independent on the selection and reproduction rates. This is not the case for M_c , whose value varies with these parameters (see Supplementary Notes). Using lattice simulations, stochastic PDE and the properties of the complex Ginzburg–Landau equation, we have derived the dependence of the critical mobility $M_c(\mu)$ on the

patterns are independent of the initial conditions. **d**, Ignoring the effects of noise, we are left with deterministic PDE that also give rise to spiralling structures. The latter share the same wavelength and frequency with those of the stochastic description but, in the absence of fluctuations, their overall size and number depend on the initial conditions and can deviate significantly from their stochastic counterparts. In **b** and **c**, the system is initially in a homogeneous state, while **d** has been generated by considering an initial local perturbation. Parameters are $\sigma = \mu = 1$ and $M = 1 \times 10^{-5}$.

reproduction rate μ (where the time-unit is set by keeping $\sigma = 1$). This enables us to analytically predict, for all values of parameters, whether biodiversity is maintained or lost. We have summarized these results in a phase diagram, reported in Fig. 4. We identify a uniform phase, in which two species go extinct (when $M > M_c(\mu)$), and a biodiverse phase (when $M < M_c(\mu)$) with coexistence of all species and propagation of spiral waves.

The generic ingredients required for the above scenario to hold are the mobility of the individuals and a cyclic dynamics exhibiting an unstable reactive fixed point. The underlying mathematical description of this class of dynamical systems is derived in terms of complex Ginzburg–Landau equations. Their universality classes reveal the robustness of the phenomena which we have reported above, that is, the existence of a critical mobility and the emergence of spiral waves; they are not restricted to specific details of the model.

Our study has direct implications for experimental research on biodiversity and pattern formation. As an example, one can envisage

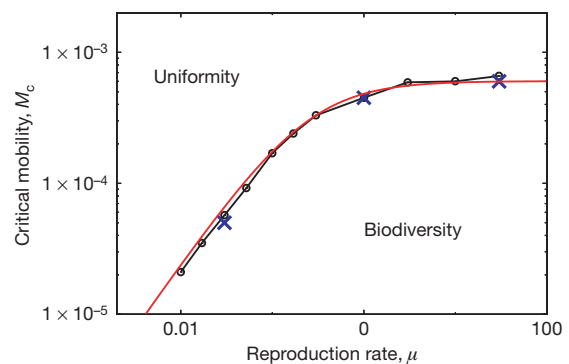


Figure 4 | Phase diagram. The critical mobility M_c as a function of the reproduction rate μ yields a phase diagram with a phase where biodiversity is maintained as well as a uniform one where two species go extinct. The time unit is set by $\sigma = 1$. On the one hand, we have computed M_c from lattice simulations, using different system sizes. The results are shown as blue crosses. On the other hand, we have calculated M_c using the approach of stochastic PDE (black dots, black lines are a guide to the eye) as well as analytically via the complex Ginzburg–Landau equation (red line). When we vary the reproduction rate, two different regimes emerge. If μ is much smaller than the selection rate, that is, $\mu \ll \sigma$, reproduction is the dominant limiter of the temporal development. In this case, there is a linear relation with the critical mobility, that is $M_c \propto \mu$, as follows from dimensional analysis. In the opposite case, if reproduction occurs much faster than selection ($\mu \gg \sigma$), the latter limits the dynamics and M_c depends linearly on σ , so that $M_c \propto \sigma$. Here, as $\sigma = 1$ is kept fixed (time-scale unit), this behaviour is reflected in the fact that M_c approaches a constant value for $\mu \gg \sigma$.

an experiment extending the study³ on colicinogenic *E. coli*. Allowing the bacteria to migrate in soft agar on a Petri dish should, for low mobilities, result in stable coexistence promoted by the formation of spiral patterns. Increasing the mobility (for example, on super-soft agar), the patterns should grow in size and finally outgrow the system at some critical value, corresponding to the threshold M_c discussed above. For even higher values of the mobility, biodiversity should be lost after a short transient time and only one species should cover the entire Petri dish. We think that the regimes of both mobilities, corresponding to the biodiverse and uniform phases, should be experimentally accessible.

We have shown how concepts from game theory combined with methods used to study pattern formation reveal the subtle influence of mobility on the temporal development of coexisting species. Many more questions and applications regarding the seminal interplay between these different fields lie ahead. As an example, it has been shown that cyclic dominance can occur in social dilemmas^{9,30}, which suggests implications of our results for the behavioural sciences.

METHODS SUMMARY

To model cyclic dominance, we use a stochastic lattice version (following work by Durrett and Levin)² of a model proposed by May and Leonard¹⁰ in 1975. As main characteristics, in the absence of spatial structure, their equations possess a deterministically unstable fixed point associated to coexistence of all three species: in the course of time, the system spirals (in the phase space) away from coexistence and moves in turn from a state with nearly only As to another one with nearly only Bs, and then to a state with nearly only Cs.

In our stochastic lattice simulations, we have arranged the three subpopulations on a two-dimensional square lattice with periodic boundary conditions. Every lattice site is occupied by an individual of species A, species B or species C, or left empty. At each simulation step, a random individual is chosen to interact with one of its four nearest neighbours: which one is also randomly determined. Whether selection, reproduction or mobility occurs, as well as the corresponding waiting time, is computed according to the reaction rates using an efficient algorithm due to Gillespie³¹. We set one generation (when every individual has reacted on average once) as the unit of time. To compute the extinction probability, we have used different system sizes, from 20×20 to 200×200 lattice sites, and sampled between 500 and 2,000 realizations. The snapshots shown in Fig. 2 result from system sizes of up to $1,000 \times 1,000$ sites.

Our stochastic PDE consist of a mobility term, nonlinear terms describing the deterministic temporal development of the nonspatial model (May–Leonard equations), and (multiplicative) white noise; see Supplementary Notes.

We have solved the resulting equations with the help of open software from the XMDS project (<http://www.xmds.org>), using the semi-implicit method in the interaction picture (SIIP) as an algorithm, spatial meshes of 200×200 to 500×500 points, and 10,000 points in the time direction.

Received 3 May; accepted 18 July 2007.

1. Durrett, R. & Levin, S. Allelopathy in spatially distributed populations. *J. Theor. Biol.* **185**, 165–171 (1997).
2. Durrett, R. & Levin, S. Spatial aspects of interspecific competition. *Theor. Pop. Biol.* **53**, 30–43 (1998).
3. Kerr, B., Riley, M. A., Feldman, M. W. & Bohannan, B. J. M. Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. *Nature* **418**, 171–174 (2002).
4. Czárán, T. L., Hoekstra, R. F. & Pagie, L. Chemical warfare between microbes promotes biodiversity. *Proc. Natl Acad. Sci. USA* **99**, 786–790 (2002).
5. Szabó, G. & Fáth, G. Evolutionary games on graphs. *Phys. Rep.* **446**, 97–216 (2007).

6. Dykhuizen, D. E. *Santa rosalia revisited: Why are there so many species of bacteria?* *Antonie Van Leeuwenhoek* **73**, 25–33 (1998).
7. Smith, J. M. *Evolution and the Theory of Games* (Cambridge Univ. Press, Cambridge, 1982).
8. Hofbauer, J. & Sigmund, K. *Evolutionary Games and Population Dynamics* (Cambridge Univ. Press, Cambridge, 1998).
9. Nowak, M. A. *Evolutionary Dynamics* (Belknap Press, Cambridge, Massachusetts, 2006).
10. May, R. M. & Leonard, W. J. Nonlinear aspects of competition between species. *SIAM J. Appl. Math.* **29**, 243–253 (1975).
11. Johnson, C. R. & Seinen, I. Selection for restraint in competitive ability in spatial competition systems. *Proc. R. Soc. Lond. B* **269**, 655–663 (2002).
12. Reichenbach, T., Mobilia, M. & Frey, E. Coexistence versus extinction in the stochastic cyclic Lotka–Volterra model. *Phys. Rev. E* **74**, 051907 (2006).
13. Jackson, J. B. C. & Buss, L. Allelopathy and spatial competition among coral reef invertebrates. *Proc. Natl Acad. Sci. USA* **72**, 5160–5163 (1975).
14. Sinervo, B. & Lively, C. M. The rock–scissors–paper game and the evolution of alternative male strategies. *Nature* **380**, 240–243 (1996).
15. Kirkup, B. C. & Riley, M. A. Antibiotic-mediated antagonism leads to a bacterial game of rock–paper–scissors *in vivo*. *Nature* **428**, 412–414 (2004).
16. Levin, S. A. Dispersion and population interactions. *Am. Nat.* **108**, 207–228 (1974).
17. Hassell, P. M., Comins, H. N. & May, R. M. Spatial structure and chaos in insect population dynamics. *Nature* **353**, 255–258 (1991).
18. Blasius, B., Huppert, A. & Stone, L. Complex dynamics and phase synchronization in spatially extended ecological systems. *Nature* **399**, 354–359 (1999).
19. King, A. A. & Hastings, A. Spatial mechanism for coexistence of species sharing a common natural enemy. *Theor. Pop. Biol.* **64**, 431–438 (2003).
20. Durrett, R. & Levin, S. The importance of being discrete (and spatial). *Theor. Pop. Biol.* **46**, 363–394 (1994).
21. Redner, S. *A Guide to First-Passage Processes* (Cambridge Univ. Press, Cambridge, 2001).
22. Hastings, A. Transients: the key to long-term ecological understanding? *Trends Ecol. Evol.* **19**, 39–45 (2004).
23. Berg, H. C. *E. coli in Motion* (Springer, New York, 2003).
24. Igoshin, O. A., Welch, R., Kaiser, D. & Oster, G. Waves and aggregation patterns in myxobacteria. *Proc. Natl Acad. Sci. USA* **101**, 4256–4261 (2004).
25. Siegert, F. & Weijer, C. J. Spiral and concentric waves organize multicellular *Dictyostelium* mounds. *Curr. Biol.* **5**, 937–943 (1995).
26. Thul, R. & Falck, M. Stability of membrane bound reactions. *Phys. Rev. Lett.* **93**, 188103 (2004).
27. Gardiner, C. W. *Handbook of Stochastic Methods* (Springer, Berlin, 1983).
28. Wiggins, S. *Introduction to Applied Nonlinear Dynamical Systems and Chaos* Ch. 2 and 3 (Springer, New York, 1990).
29. Aranson, I. S. & Kramer, L. The world of the complex Ginzburg–Landau equation. *Rev. Mod. Phys.* **74**, 99–143 (2002).
30. Hauert, C., de Monte, S., Hofbauer, J. & Sigmund, K. Volunteering as red queen mechanism for cooperation in public goods games. *Science* **296**, 1129–1132 (2002).
31. Gillespie, D. T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **22**, 403–434 (1976).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Bathe and M. Leisner for discussions on the manuscript. Financial support of the German Excellence Initiative via the program “Nanosystems Initiative Munich (NIM)” as well as the SFB “Manipulation of Matter at the Nanometer Length Scale” is gratefully acknowledged. M.M. is grateful to the Alexander von Humboldt Foundation for support through a fellowship.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.F. (frey@lmu.de).

LETTERS

A sequence-based variation map of 8.27 million SNPs in inbred mouse strains

Kelly A. Frazer¹, Eleazar Eskin², Hyun Min Kang³, Molly A. Bogue⁴, David A. Hinds¹, Erica J. Beilharz¹, Robert V. Gupta¹, Julie Montgomery¹, Matt M. Morenson¹, Geoffrey B. Nilsen¹, Charit L. Pethiyagoda¹, Laura L. Stuve¹, Frank M. Johnson⁵, Mark J. Daly^{6,7}, Claire M. Wade^{6,7} & David R. Cox¹

A dense map of genetic variation in the laboratory mouse genome will provide insights into the evolutionary history of the species¹ and lead to an improved understanding of the relationship between inter-strain genotypic and phenotypic differences. Here we resequence the genomes of four wild-derived and eleven classical strains. We identify 8.27 million high-quality single nucleotide polymorphisms (SNPs) densely distributed across the genome, and determine the locations of the high (divergent subspecies ancestry) and low (common subspecies ancestry) SNP-rate intervals^{2–6} for every pairwise combination of classical strains. Using these data, we generate a genome-wide haplotype map containing 40,898 segments, each with an average of three distinct ancestral haplotypes. For the haplotypes in the classical strains that are unequivocally assigned ancestry, the genetic contributions of the *Mus musculus* subspecies—*M. m. domesticus*, *M. m. musculus*, *M. m. castaneus* and the hybrid *M. m. molossinus*—are 68%, 6%, 3% and 10%, respectively; the remaining 13% of haplotypes are of unknown ancestral origin. The considerable regional redundancy of the SNP data will facilitate imputation of the majority of these genotypes in less-densely typed classical inbred strains to provide a complete view of variation in additional strains.

Classical strains are derived from a limited number of founding mice that were genetically diverse owing to the mixture of subspecies of the genus *Mus*^{2,3}: *M. m. musculus*, *M. m. castaneus*, *M. m. domesticus* and the hybrid *M. m. molossinus* (Fig. 1). In contrast, most of the wild-derived strains used in research are largely genetically pure because they are derived from mice captured in the wild, which were inbred to homozygosity^{4,5}. The unique breeding history of classical strains is reflected in their genomic structure; recent pairwise sequence comparisons reveal alternating intervals of low and high sequence-variation, representing segments of the genome where the two strains have either common (for example, both *domesticus*) or divergent (for example, one *domesticus* and one *musculus*) subspecies ancestry, respectively^{6–10}. The locations and boundaries of the genomic intervals composed of common and divergent ancestry sequences are different for each pair of classical strains and need to be empirically determined.

Resequencing methods determine the DNA sequences of individual members of a species for which a high-quality reference genome is available more efficiently than *de novo* sequencing. Here, we resequenced the genomes of the 15 mouse strains (Box 1), using oligonucleotide arrays⁶ designed to query ~ 1.49 billion bases (58%) of the 2.57-billion-basepair C57BL/6J reference genome¹¹ (NCBI

Build 36) for all possible single-base substitutions¹² (Supplementary Fig. 1). These 1.49 billion bases represent primarily unique sequences, but also include some lightly masked repetitive sequence (Supplementary Table 1). Nucleotide variation from the C57BL/6J reference sequence was identified by hybridizing labelled amplified DNA from the 15 strains to the arrays. The arrays were then scanned and the feature intensity data analysed with pattern recognition base-calling and SNP-calling algorithms^{13,14}. To make the resequencing data available to individual researchers we developed a method for representing the array data as sequence traces analogous to conventional dideoxy sequencing (Supplementary Fig. 2).

We identified 8,266,653 SNPs with unique positions in the mouse genome (NCBI Build 36), the vast majority of which are novel. Of the SNP submissions clustered in dbSNP Build 126, 92.5% were either the only instance of the associated cluster or the member of the cluster submitted first. We assessed the quality of our SNP data, determining that the false-positive rate of discovery is 2% and the accuracy of our genotype calls is greater than 99% (Supplementary Information). To assess our false-negative rate for SNP discovery we selected all SNPs on chromosome 4 where the A/J strain in the Celera WGS resequencing data¹⁵ had a different allele from the C57BL/6J sequence and for which the same genotypes were validated by an independent study at the Broad Institute (unpublished data). This resulted in a set of 5,553 validated SNPs on chromosome 4, of which 3,936 SNPs (70.9%) were in sequences tiled on the arrays and 2,366 (42.6%) had the alternate allele detected in at least one of the 15 strains analysed by array-based resequencing. From this analysis we estimate that 2,366 out of 3,939, or ~60%, of SNPs in tiled sequences and ~43% of all SNPs present in the classical strains were identified in our array-based resequencing study.

Of the 8.27 million SNPs identified among all strains, 3.39 million (41%) were polymorphic when only the 11 classical strains were considered and 6.46 million (78%) were polymorphic when only the 4 wild-derived strains were considered. Given that we discovered 3.4 million SNPs in the 12 classical strains and have a 43% SNP discovery rate, we estimate that in total there are ~8 million SNPs in the classical strains. Interestingly, 1.81 million (22%) of the SNPs polymorphic among the classical strains were monomorphic in the wild-derived strains, indicating that a substantial amount of un-sampled ancestral genotypes contributed to the generation of the classical strains.

With a total of 2.57 billion base pairs of available mouse genome reference sequence, the 8.27 million SNPs identified among all strains yields an average density of 1 SNP per 311 base pairs (Supplementary

¹Perlegen Sciences, 2021 Stierlin Court, Mountain View, California 94043, USA. ²Department of Computer Science and Department of Human Genetics, University of California, Los Angeles, Los Angeles, California 90095, USA. ³Department of Computer Science and Engineering, University of California, San Diego, La Jolla, California 92093, USA. ⁴The Jackson Laboratory, Bar Harbor, Maine 04609, USA. ⁵Toxicology Operations Branch, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709, USA. ⁶Broad Institute of Harvard and MIT, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. ⁷Center for Human Genetic Research, Massachusetts General Hospital, 185 Cambridge St, Boston, Massachusetts 02114, USA.

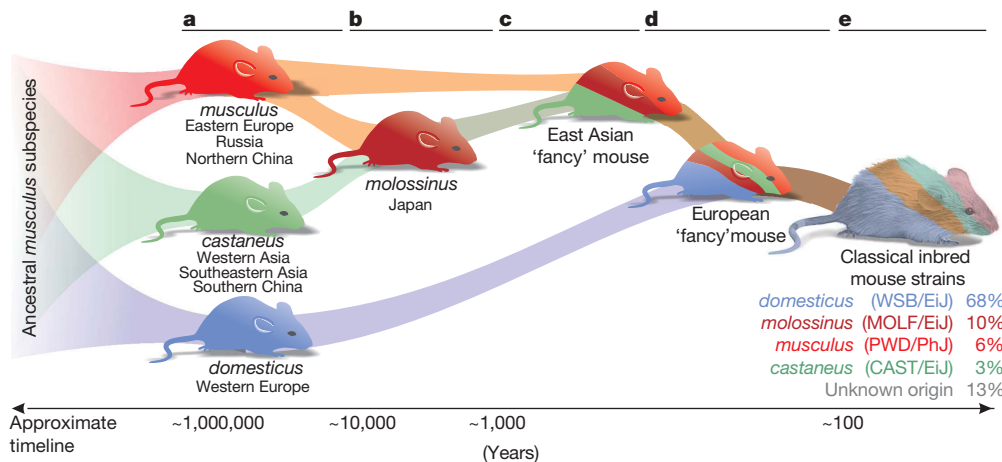


Figure 1 | Origin of modern classical strains. **a**, The subspecies *musculus*, *castaneus* and *domesticus* subspecies diverged from a common ancestor about one million years ago. **b**, The subspecies *molossinus* is a more recent hybrid between *musculus* and *castaneus*¹⁶. **c**, In the eighteenth century, mouse fanciers in Japan and China inter-bred the Asian *Mus musculus* subspecies to produce varieties of mice with different coat colours and

behavioural characteristics as pets^{2,16}. **d**, In Victorian England, 'fancy' mice were imported from Asia and bred with local mice. In the early twentieth century, mating programmes using a limited number of founding 'fancy' mice were started in the USA, giving rise to many of the modern classical strains². **e**, The relative contributions of the *Mus musculus* subspecies to haplotypes in classical strains.

Table 2); for the subset of SNPs polymorphic only in the wild-derived strains, the density is 1 SNP every 397 base pairs; and for the subset of SNPs polymorphic only among the classical strains, 1 SNP every 756 base pairs. Given our 43% SNP discovery rate, these data agree with previous estimates of classical strain diversity of 1 SNP every 250–300 base pairs⁷.

To investigate the uniformity of genome coverage of our SNP set, we examined the distribution of inter-SNP distances. For the complete SNP set, 95% of the inter-SNP intervals are 1 kb or less, and 99.7% are 10 kb or shorter (Supplementary Fig. 3a). Considering the wild-derived strains and the classical strains separately, the equivalent values are 93% and 99.5% for wild-derived strains, and 87% and 98.7% for classical strains. Thus, the vast majority of inter-SNP distances are shorter than 1 kb. We also examined the fraction of the available genome within a given maximum distance of a typed SNP (Supplementary Fig. 3b). For our complete SNP set, 75% of the genome was within 1 kb of an SNP, and 93% of the genome was within 5 kb of an SNP. For SNPs polymorphic in the wild-derived inbred strains, 70% and 91% of the genome was within 1 kb and 5 kb of an SNP, respectively. For SNPs polymorphic in the classical strains,

53% and 81% of the genome was within 1 kb and 5 kb of an SNP, respectively. The inter-SNP interval and maximum distance analyses provide complementary information, with both indicating that the genome coverage of our 8.27 million SNP set is dense and uniform.

We next determined the common (low SNP-rate) and divergent (high SNP-rate) ancestral intervals across the genome for each pairwise combination of the classical strains, using a dynamic programming algorithm to identify regions of shared ancestry among the strains. We included the genotypes of the reference strain C57BL/6J so in total there were 12 strains analysed. For the 66 pairwise comparisons, there was an average of 5,519 low SNP-rate (1 SNP per 21 kb) and/or high SNP-rate (1 SNP per 440 base pairs) intervals across the genome (Supplementary Table 3). The low SNP-rate intervals contained 62% of the bases and extended for an average of 568 kb, whereas the high SNP-rate intervals contained fewer bases (38%) and were shorter, at 355 kb.

To generate a haplotype map across the genome, we mapped all the ancestry breakpoints in the 12 classical strains, defined as locations at which any of the pairwise comparisons transitioned from a low SNP-rate to a high SNP-rate interval or vice versa. The regions between these breakpoints represent segments in which all strains have a single unbroken ancestral haplotype (Fig. 2). Our genome-wide map consists of 40,898 segments with an average length of 58 kb (ranging from 1 kb to 3 Mb), covering approximately 90% of the mouse genome (Supplementary Table 4). On average there are 3.05 ± 0.98 different haplotypes observed per segment in the 12 classical strains.

In each pairwise comparison the low SNP-rate intervals extend through approximately 10 segment boundaries. To characterize the total amount of genetic variation in a region of a given size, we consider strain distribution patterns (SDPs) or distinct groupings of the strains that are based on SNPs within a region (Supplementary Table 5). The number of SDPs represents the number of non-redundant SNPs within a defined region (that is, the number of SNPs required to capture the variation content of that region). Less than 2% of the SNPs in a 50-kb region represent singleton SDPs (and less than 0.5% of the SNPs in a 5-Mb window). The number of SDPs accounting for 95% of all patterns observed is 3.3 in 50-kb windows and only 28.4 in 5-Mb windows. Thus, although slightly fewer than half of all SNPs in the classical strains were discovered, the redundancy demonstrated in the SDP analysis indicates that very few distinct variation patterns are uncaptured by this data set.

The inclusion of four wild-derived strains in our study allowed us to determine the ancestry of many of the haplotypes in the classical

Box 1 | Fifteen mouse strains selected for resequencing

The following criteria were considered in the selection process: widespread use in research; availability of ancillary genetic resources such as recombinant inbred and/or consomic lines¹⁸; inclusion in the Mouse Phenome Project^{19–21}; maximization of phenotypic and genetic diversity; and ease of breeding to enable follow-up studies. The 15 selected strains can be divided into three groups:

- Group one consists of eight classical strains commonly used for research, of which several have served as progenitors of recombinant inbred strains and/or are part of consomic panels with C57BL/6J (www.jax.org/); these are: 129S1/SvImJ, A/J, AKR/J, BALB/cByJ, C3H/HeJ, DBA/2J, FVB/NJ, NOD/LtJ.
- Group two consists of three classical strains that are useful models of complex human diseases: BTBR T+tf/J, which is used as a model to study behavioural abnormalities²² and autism²³; KK/HIJ, which was chosen because of its distinct lineage from the other strains in the study¹ and its use as a model for type II diabetes²⁴, and NZW/LacJ, which is used as a model to study autoimmune diseases.
- Group three consists of four wild-derived strains: CAST/EiJ (*castaneus*), MOLF/EiJ (*molossinus*), PWD/PhJ (*musculus*) and WSB/EiJ (*domesticus*). Because the origins of the classical strains stem from the interbreeding of the *Mus musculus* subspecies (Fig. 1), this group provides important information on ancestral sequence diversity.

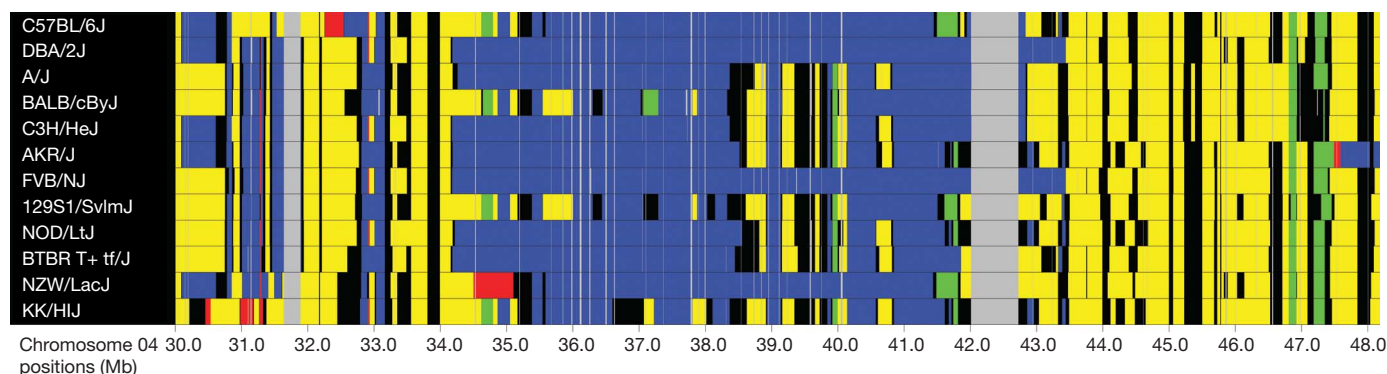


Figure 2 | Ancestral haplotype map for an 18-Mb region on chromosome 4. Colours represent an ancestral origin of WSB/EiJ (yellow), PWD/PhJ (red), CAST/EiJ (green) or MOLF/EiJ (blue) strains. For the 12 classical inbred

strains, each segment is coloured to indicate its predicted ancestral origin. The grey colour represents gaps in the SNP coverage and the black colour represents either ambiguous or unknown ancestral origin.

strains. A dynamic programming algorithm assigned the most-likely wild-derived ancestor for each classical strain in each segment and mapped these predictions to the ancestry breakpoints. On average, ancestry for 67% of the genome of each classical strain could be unequivocally assigned to one of the *Mus* subspecies. Ancestry for 10% of the genome was unequivocally different from all four wild-derived strains. The remaining portion of the genome has ambiguous ancestry because either the region matched two or more of the wild-derived strains (3%) or there was insufficient coverage to determine ancestry (20%) because the algorithm required all 4 wild-derived strains to be successfully genotyped. For the classical strain haplotypes that could be assigned ancestry, the *domesticus* strain (WSB/EiJ) matched 68%, the *molossinus* strain (MOLF/EiJ) 10%, the *musculus* strain (PWD/PhJ) 6%, and *castaneus* (CAST/EiJ) 3%, with the remaining 13% unequivocally ancestrally divergent from the four wild-derived strains (Fig. 1, Supplementary Table 6). Therefore, of the *Mus* subspecies, *domesticus* made the largest genetic contribution to the classical strains, a finding consistent with previous reports that were based on candidate region analyses^{7,16}. Applying a similar analysis to the hybrid *molossinus* strain we identify its ancestry as 57% *musculus*, 12% *castaneus*, 7% *domesticus* and 24% of unknown ancestry, which is also consistent with previous reports.

Interestingly, the ancestral contribution represented by each of the four wild-derived strains greatly differs across the genome even at the level of entire chromosomes (Table 1). In the regions with unambiguous ancestry, *domesticus* contribution ranges from 80%

on chromosome 18 to only 51% on chromosome 14. The equivalent range for *molossinus* is from 22% on chromosome 4 to only 1% on chromosome 16, for *musculus* from 28% on chromosome 14 to only 1% on chromosome 5, and for *castaneus* from 8% on the chromosome 19 to less than 1% on chromosome 18. Although the relative ancestral contributions of the 4 wild-derived strains to each of the 12 classical strains is highly similar at the genome-wide level, the distribution of ancestral origins varies greatly at the chromosome level (Supplementary Table 7). For example, the *molossinus* contribution on chromosome 7 is 2% in C57BL/6J and 19% in DBA/2J, whereas on chromosome 12 the *molossinus* contribution is 14% in C57BL/6J and 1% in DBA/2J.

Our data are accessible through NCBI databases, with SNPs deposited in dbSNP and trace files in the Trace Archive. In addition, we have established a website (<http://mouse.perlegen.com>), which researchers can use to download SNPs, genotypes and LR-PCR primer pairs, all of which are currently mapped to NCBI Build 36. The site hosts a genome browser¹⁷ (Supplementary Fig. 4)—which allows users to visualize, for specific genomic intervals, the locations of SNPs, LR-PCR primer pairs, and intervals of non-repetitive sequence that were assayed in this study—and a webserver, where researchers can browse and download the ancestral haplotype map to obtain ancestry information for any region of the mouse genome.

We anticipate that data generated from SNP sets that are less dense but genotyped in strains not included in our study, such as the Broad Institute resource of 150,000 genome-wide genotypes on the most commonly used 100 classical strains, will be used to map additional strains to our ancestral haplotype map. This will allow genotypes for most of our 8.27 million SNPs to be imputed for these unsequenced strains on the basis of their shared ancestral haplotypes.

Table 1 | Ancestral contribution of each of the four wild-derived strains and an unknown strain, which is unequivocally ancestrally divergent from any of the four wild-derived strains, for each chromosome

| Chromosome | WSB/EiJ | PWD/PhJ | CAST/EiJ | MOLF/EiJ | Unknown |
|-------------|---------|---------|----------|----------|---------|
| 01 | 0.6684 | 0.0818 | 0.0202 | 0.0930 | 0.1367 |
| 02 | 0.6337 | 0.0267 | 0.0542 | 0.1935 | 0.0919 |
| 03 | 0.7054 | 0.0261 | 0.0463 | 0.0763 | 0.1458 |
| 04 | 0.6066 | 0.0554 | 0.0136 | 0.2227 | 0.1018 |
| 05 | 0.6296 | 0.0082 | 0.0122 | 0.2215 | 0.1286 |
| 06 | 0.6384 | 0.0787 | 0.0286 | 0.1005 | 0.1538 |
| 07 | 0.7348 | 0.0309 | 0.0103 | 0.0761 | 0.1479 |
| 08 | 0.7166 | 0.0373 | 0.0061 | 0.0756 | 0.1644 |
| 09 | 0.5613 | 0.1763 | 0.0288 | 0.0996 | 0.1340 |
| 10 | 0.6799 | 0.0260 | 0.0160 | 0.1023 | 0.1758 |
| 11 | 0.6907 | 0.0196 | 0.0158 | 0.1749 | 0.0990 |
| 12 | 0.7193 | 0.0200 | 0.0677 | 0.0415 | 0.1516 |
| 13 | 0.7113 | 0.0688 | 0.0143 | 0.0887 | 0.1169 |
| 14 | 0.5107 | 0.2805 | 0.0099 | 0.0985 | 0.1005 |
| 15 | 0.7166 | 0.0465 | 0.0403 | 0.0720 | 0.1246 |
| 16 | 0.7916 | 0.0095 | 0.0347 | 0.0148 | 0.1495 |
| 17 | 0.6725 | 0.0494 | 0.0543 | 0.0348 | 0.1890 |
| 18 | 0.7999 | 0.0218 | 0.0033 | 0.0273 | 0.1478 |
| 19 | 0.7537 | 0.0105 | 0.0776 | 0.0525 | 0.1057 |
| X | 0.8712 | 0.0179 | 0.0811 | 0.0145 | 0.0154 |
| Genome-wide | 0.6846 | 0.0562 | 0.0304 | 0.1002 | 0.1287 |

METHODS SUMMARY

Array design. The 1.49 million bases tiled on the arrays included at least a portion of 27,915 (97%) of the 28,874 annotated genes in the mouse genome (NCBI Build 36.1), and at least a portion of 201,195 (98%) of the 206,338 annotated exons. To confirm the identity of the strains processed in each hybridization experiment, on all arrays we included a set of 96 SNPs distributed across 8 different genome regions, the genotypes of which allowed unambiguous discrimination between the 15 inbred strains (Supplementary Table 8).

Samples. Resequencing was performed using DNA isolated from pedigree breeding stock animals (Jackson Laboratory (<http://www.jax.org/>)). To prepare the DNA samples for analysis, we designed 241,806 long-range PCR (LR-PCR) primer pairs to generate amplicons ranging from 3 to 12 kb in length. Together these primer pairs were used to amplify essentially the entire genomes of each of the 15 strains.

Resequencing. High-density oligonucleotide array technology was used as described^{13,14}. Nucleotide variation from the C57BL/6J reference sequence was identified by hybridizing labelled amplified DNA from the 15 strains to the arrays, which were then scanned and the feature intensity data analysed with pattern recognition base-calling and SNP-calling algorithms^{13,14}.

Ancestral haplotype map. To partition the genomes into segments in our pairwise comparisons of the 12 classical strains, we constructed a hidden Markov model with two states for each SNP position (one for high and one for low SNP-rate) and applied the Viterbi algorithm to assign each SNP to one of the two segment types. We constructed the haplotype map by defining a block boundary at each position where a segment transitioned from a low SNP-rate to a high SNP-rate segment or vice versa in one of the 66 pairwise comparisons. To identify ancestral origins of haplotypes, we constructed a hidden Markov model with five states for every SNP, with four states each representing one of the four wild-derived strains and one state representing an unknown ancestral origin, and applied the forward-backward algorithm to assign the ancestry to each SNP in each of the classical strains.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 14 April 2007; accepted 5 July 2007.

Published online 29 July 2007.

- Beck, J. A. *et al.* Genealogies of mouse inbred strains. *Nature Genet.* **24**, 23–25 (2000).
- Wade, C. M. & Daly, M. J. Genetic variation in laboratory mice. *Nature Genet.* **37**, 1175–1180 (2005).
- Bonhomme, F., Guénet, J.-L., Dod, B., Moriwaki, K. & Bulfield, G. The polyphyletic origin of laboratory inbred mice and their rate of evolution. *Biol. J. Linnean Soc.* **30**, 51–58 (1987).
- Petkov, P. M. *et al.* An efficient SNP system for mouse genome scanning and elucidating strain relationships. *Genome Res.* **14**, 1806–1811 (2004).
- Ideraabdullah, F. Y. *et al.* Genetic and haplotype diversity among wild-derived mouse inbred strains. *Genome Res.* **14**, 1880–1887 (2004).
- Frazer, K. A. *et al.* Segmental phylogenetic relationships of inbred mouse strains revealed by fine-scale analysis of sequence variation across 4.6 Mb of mouse genome. *Genome Res.* **14**, 1493–1500 (2004).
- Wade, C. M. *et al.* The mosaic structure of variation in the laboratory mouse genome. *Nature* **420**, 574–578 (2002).
- Wiltshire, T. *et al.* Genome-wide single-nucleotide polymorphism analysis defines haplotype patterns in mouse. *Proc. Natl Acad. Sci. USA* **100**, 3380–3385 (2003).
- Yalcin, B. *et al.* Unexpected complexity in the haplotypes of commonly used inbred strains of laboratory mice. *Proc. Natl Acad. Sci. USA* **101**, 9734–9739 (2004).
- Zhang, J. *et al.* A high-resolution multistrain haplotype analysis of laboratory mouse genome reveals three distinctive genetic variation patterns. *Genome Res.* **15**, 241–249 (2005).
- Waterston, R. H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
- Chee, M. *et al.* Accessing genetic information with high-density DNA arrays. *Science* **274**, 610–614 (1996).
- Hinds, D. A. *et al.* Whole-genome patterns of common DNA variation in three human populations. *Science* **307**, 1072–1079 (2005).
- Patil, N. *et al.* Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* **294**, 1719–1723 (2001).
- Mural, R. J. *et al.* A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome. *Science* **296**, 1661–1671 (2002).
- Silver, L. M. *Mouse Genetics: Concepts and Applications* (Oxford University Press, New York, 1995).
- Stein, L. D. *et al.* The generic genome browser: a building block for a model organism system database. *Genome Res.* **12**, 1599–1610 (2002).
- Singer, J. B. *et al.* Genetic dissection of complex traits with chromosome substitution strains of mice. *Science* **304**, 445–448 (2004).
- Grubb, S. C., Churchill, G. A. & Bogue, M. A. A collaborative database of inbred mouse strain characteristics. *Bioinformatics* **20**, 2857–2859 (2004).
- Bogue, M. A. & Grubb, S. C. The Mouse Phenome Project. *Genetica* **122**, 71–74 (2004).
- Bogue, M. Mouse Phenome Project: understanding human biology through mouse genetics and genomics. *J. Appl. Physiol.* **95**, 1335–1337 (2003).
- Wahlsten, D., Metten, P. & Crabbe, J. C. Survey of 21 inbred mouse strains in two laboratories reveals that BTBR T/+^{tf/tf} has severely reduced hippocampal commissure and absent corpus callosum. *Brain Res.* **971**, 47–54 (2003).
- Moy, S. S. *et al.* Mouse behavioral tasks relevant to autism: phenotypes of 10 inbred strains. *Behav. Brain Res.* **176**, 4–20 (2007).
- Nakamura, M. A diabetic strain of the mouse. *Proc. Jap. Acad.* **38**, 348–352 (1962).
- Ewing, B. & Green, P. Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Res.* **8**, 186–194 (1998).
- Viterbi, A. J. Error bounds for convolutional codes and an asymptotically optimal decoding algorithm. *IEEE Trans. Information Theory* **13**, 260–269 (1967).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Work was supported by funding from the NIEHS. H.M.K. and E.E. are partially supported by the NSF. H.M.K. is partially supported by a Samsung Scholarship. E.E. is partially supported by the NIH. At Perlegen Sciences, we thank A. Kloek for assistance with manuscript preparation; B. Nguyen, X. Chen, P. Chu, R. Patel, P.-E. Jiao, R. Irikat and J. Kwon for assistance with DNA sample preparation and hybridization of the high-density oligonucleotide arrays; R. Vergara for primer handling; H. Huang and W. Barrett for designing the high-density arrays; T. Genschoreck and J. Sheehan for data quality control; and S. Osborn for assistance with website development and data delivery. At NIEHS, we thank D. A. Schwartz, K. Olden, S. Wilson, L. Birnbaumer, J. Bucher, W. T. Schrader and D. M. Klotz for constructive scientific discussions, and J. A. Lewis and T. Hardee for administrative support. At The Jackson Laboratory, we thank S. Deveau and JAX DNA Resources for DNA sample preparation.

Author Contributions L.L.S., J.M., C.L.P. and K.A.F. supervised the experiments. K.A.F., D.R.C., M.J.D., F.M.J., E.J.B. and M.A.B. designed the study. H.M.K., E.E., C.M.W., D.A.H., G.B.N., R.V.G. and M.M.M. performed data analysis. K.A.F., with help from E.J.B., D.A.H., E.E. and C.M.W., wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests; details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.A.F. (frazer.kelly@scrippshealth.org).

METHODS

High-density oligonucleotide array design. Our study was initiated before the C57BL/6J finished reference sequence was available for the entire mouse genome. Thus, multiple NCBI genome assemblies were used for the design of resequencing arrays for different parts of the genome: Build 33 for chromosomes 2, 4, 11 and X; Build 34 for finished sequences greater than 315 kb from chromosomes 1, 3, 5–10 and 12–19; Build 35 for finished sequences greater than 315 kb from chromosomes 1, 3, 5–10 and 12–19 that were not already covered; and Build 36 for remaining sequence, including chromosome Y and contigs not on the main assembly.

The C57BL/6J reference sequence was screened and annotated for interspersed repeats and low-complexity DNA sequences by RepeatMasker using the $-s$ sensitive setting and the $-m$ flag to specify murine sequence. Additionally, the sequence was processed using the Tandem Repeat Finder program, masking out repeats with a period of 12 or less. The masked sequence was downloaded from the UCSC Genome Bioinformatics website (<http://genome.ucsc.edu/>). Of the 2,567,283,971 bases from Build 36, approximately 42.3% are masked as repetitive. Although we primarily tiled the unique non-masked sequences for a fraction of the genome, we used the $-\text{div } 10$ option in Repeat Masker (<http://www.repeatmasker.org/>), which only masks sequences less than 10% diverged from the canonical repeats in the database. For 14,806 intervals distributed across the 19 autosomes, an additional 130,236,585 bp of lightly masked sequence using the $-\text{div } 10$ option and 10,633,723 bp of unmasked sequence (an aggregate total of 140,900,308 bp resulting from different masking options) were tiled on the arrays (Supplementary Table 1). A total of 1,488,604,099 bp of unique or lightly masked repetitive sequence was used to design high-density oligonucleotide arrays.

The arrays were designed such that each of the 1,488,604,099 nucleotides was interrogated by eight 25-mer oligonucleotides (features) that were synthesized and attached to a glass surface (Supplementary Fig. 1). Each set of eight features consisted of four identical to the forward-strand reference sequence from position -12 to $+12$ with respect to the base to be queried (position 0), with position 0 represented by each of the four bases A, C, G and T. The remaining four features were similarly designed for the reverse strand. The arrays were synthesized by Affymetrix using light-directed photolithography in conjunction with chemical coupling to direct the synthesis of the 25-mer oligonucleotides. For this study, we used large arrays containing approximately 180 million $7\text{ }\mu\text{m}$ features each. With each array able to assay around 22 Mb of unique genomic sequence, we used 68 array designs to cover the entire mouse genome, excluding repetitive sequences. **Sample preparation.** Using the C57BL/6J reference sequence and the Oligo 6 program we designed 241,806 LR-PCR primer pairs with amplicons ranging from 3 to 12 kb in length, with an average size of 10,336 bp. The primers were typically between 28 and 32 nucleotides in length with a melting temperature of $>65^\circ\text{C}$.

Resequencing was performed using DNA samples isolated from between four and fourteen individual male mice per strain (Supplementary Table 9). PCR reactions (1-plex or 2-plex) were performed as follows (per reaction): 30 ng of genomic DNA from one of the 15 inbred mouse strains was amplified using 1.7 μM of each LR-PCR primer, 0.6 U MasterAmp extra long Taq polymerase (Epicentre Technologies), 20 ng ml^{-1} TaqStart antibody (Clontech), 0.1 \times TaqStart Antibody buffer (Clontech), 0.4 mM dNTPs, 23 mM Tricine, 3% DMSO, 45 mM Trizma, 2.7 mM MgCl_2 , 12.6 mM $(\text{NH}_4)_2\text{SO}_4$, 2.6 mM KCl and 0.4 \times MasterAmp PCR Enhancer with Betaine (Epicentre Technologies), in a volume of 12 μl . The reactions were performed using a Perkin-Elmer 9700 thermocycler as follows: initial denaturation for 3 min at 95°C ; 10 cycles at 94°C for 2 s and 64°C for 15 min; 28 cycles of 94°C for 2 s and 64°C for 15 min, with a 20 s increase per cycle; and a final extension of 60 min at 62°C .

Amplicons to be hybridized together on the array ($\sim 11\text{--}14\text{ Mb}$) were combined into one tube, purified using Centricon Plus-20 filter devices (Millipore), quantified using optical density spectrophotometry, and normalized to a concentration of $2\text{ }\mu\text{g }\mu\text{l}^{-1}$. The amplicons were then fragmented to a peak fragment size of 100 bp by incubation of 200 μg of purified PCR product for 30 min at 37°C in 1 \times One-Phor-All Buffer PLUS (Amersham) and 0.001 U μl^{-1} amplification grade DNase I (Invitrogen). The DNase I was then heat-inactivated at 99°C for 10 min. The fragmentation reactions were then labelled with either biotin or fluorescein for 90 min at 37°C in 1 \times One-Phor-All Buffer PLUS, 13600 U recombinant TdT (Roche Applied Science), and either 0.1 mM each of biotin-16-ddUTP and biotin-16-dUTP (Roche Applied Science) or 0.1 mM each of fluorescein-12-ddUTP and fluorescein-12-dUTP (PerkinElmer). A subset of the samples (for 2 out of 68 arrays) were labelled using the same protocol with either biotin as described above or 0.1 mM each of digoxigenin-11-ddUTP and digoxigenin-11-dUTP (Roche Applied Science).

Array hybridization. The arrays each containing $\sim 22\text{ Mb}$ of tiled DNA sequence were physically segmented into three chambers and each chamber

was hybridized with a different DNA/hybridization mixture containing labelled target DNA at $\sim 11\text{--}14\text{ Mb}$ complexity. The labelled target DNA was prepared for hybridization by combining biotin-labelled amplicons from one strain and fluorescein- or digoxigenin-labelled amplicons from another strain, and adding 40 μl of 10 mg ml^{-1} yeast RNA (Applied Biosystems) and 200 μl 10 mg ml^{-1} herring sperm DNA (Promega), for a total of 640 μl . Combining the DNA from the two different strains labelled with distinct fluorophores in a single array hybridization allowed us to use half the number of arrays otherwise needed (8 copies of each of the 68 individual arrays spanning the genome versus 15 copies of each of these arrays). The 640 μl labelled target DNA was denatured for 10 min at between 90°C and 100°C and then snap cooled at 4°C . It was then added to 4.25 ml of hybridization buffer plus 125 μl 20 mg ml^{-1} BSA, to give a final solution of: 0.08 mg ml^{-1} yeast RNA, 0. mg ml^{-1} herring sperm DNA, 0.5 mg ml^{-1} BSA, 2.94 M TMACl, 0.01 M Tris (pH 8.0), 0.01% Triton X-100, 0.05 nM b-948 biotin control oligo (Sigma-Aldrich), and 0.05 nM fl-948 fluorescein control oligo (Sigma-Aldrich). Hybridization of the target DNA to the microarrays took place at 50°C for 17 h with constant rotation.

After hybridization, the arrays were rinsed with 1 \times MES and stained for detection of the biotin- and fluorescein-labelled hybridized targets by 30 min incubations at room temperature with the following series of 5 stain solutions (each of which was in 1 \times MES, 0.1% Triton X-100, and 2.5 mg ml^{-1} BSA): Stain 1 with 5 $\mu\text{g }\mu\text{l}^{-1}$ streptavidin (Invitrogen) plus 10 $\mu\text{g }\mu\text{l}^{-1}$ Alexa 488-rabbit anti-fluorescein antibody (Invitrogen); Stain 2 with 1.25 $\mu\text{g }\mu\text{l}^{-1}$ biotinylated anti-streptavidin (Vector Laboratories); and Stain 3 with 1 $\mu\text{g }\mu\text{l}^{-1}$ streptavidin-phycoerythrin-Cy5 (BD Biosciences) plus 10 $\mu\text{g }\mu\text{l}^{-1}$ Alexa 488 goat anti-rabbit antibody (Invitrogen). To enhance the fluorescein signal, the arrays were then incubated with Stain 4, containing 10 $\mu\text{g }\mu\text{l}^{-1}$ Alexa 488-rabbit anti-fluorescein antibody, and Stain 5, with 10 $\mu\text{g }\mu\text{l}^{-1}$ Alexa 488 goat anti-rabbit antibody, each for 30 min. The arrays were rinsed with 1 \times MES between stain incubations, and washed at high stringency in 0.2 \times SSPE, 0.01% TX-100 for 60 min at 37°C after the completion of staining. The arrays were then scanned using custom-built confocal scanners.

For the arrays using the biotin and digoxigenin labelling strategy, the labelled targets were detected by 20 min incubations at room temperature using the following series of 4 stain reagents (in 1 \times MES, 0.1% Triton X-100, and 2.5 mg ml^{-1} BSA): Stain 1 with 2.5 $\mu\text{g }\mu\text{l}^{-1}$ Anti-Digoxigenin Ab, clone 1.71.256, mouse IgG1 (Roche Applied Science) plus 5 $\mu\text{g }\mu\text{l}^{-1}$ streptavidin (Invitrogen); Stain 2 with 5 $\mu\text{g }\mu\text{l}^{-1}$ anti-streptavidin (Rabbit) Biotin Conjugated (Rockland Immunochemicals); Stain 3 with 5 $\mu\text{g }\mu\text{l}^{-1}$ Alexa Fluor 647-R-phycoerythrin goat anti-mouse IgG (Invitrogen) plus 1 $\mu\text{g }\mu\text{l}^{-1}$ streptavidin, Alexa Fluor 488 conjugate; and Stain 4 with 9 $\mu\text{g }\mu\text{l}^{-1}$ AffiniPure Mouse Anti-Goat IgG (H+L). To enhance the digoxigenin signal, the arrays were then incubated further with Stain 3 followed by Stain 4. The arrays were rinsed with 1 \times MES between incubations, and washed at high stringency in 0.2 \times SSPE, 0.01% TX-100 for 60 min at 37°C after the completion of staining. The arrays were then scanned using custom-built confocal scanners.

Base-calling algorithm. Fluorescence intensity data from an oligonucleotide array scan were first processed to determine an average intensity for each feature on the array. This yielded 8 data points per sequence position: one each for A, C, G and T, for forward and reverse strands. We determined a direct sequence for each strand, consisting of the brightest base at each position (Supplementary Fig. 1). At each position, we also determined the local 'conformance' of the array data, as the fraction of base calls that match the reference sequence within a sliding window. For a position where the direct call matches the reference base, this window consists of bases at positions -10 to $+10$. In the immediate vicinity of an alternative base call, hybridization intensities are reduced owing to the presence of a one-base mismatch base between the target and probe DNA. To avoid the reduced-intensity interval in these cases, we alter the window to span bases -20 to -10 , and $+10$ to $+20$.

A strict base call is made for a sequence position when the ratio of the brightest to next-brightest feature is greater than a threshold of 1.3 for biotin-labelled DNA and 1.1 for fluorescein-labelled DNA, and the conformance around that position is at least 0.80 on both strands. A relaxed base call is made if these criteria are met for just one strand and the other strand is ambiguous (that is, it did not pass either the intensity ratio or conformance requirements). For alternate base calls that do not match the reference sequence, we also require that there are no brighter alternate calls meeting these criteria within positions -5 to $+5$.

SNP detection algorithm. For polymorphism detection, we created a consensus sequence of strict base calls that are confirmed on both strands. Again, alternate consensus calls are excluded if there is a brighter (average intensity over both strands) alternate consensus call within positions -5 to $+5$. Putative polymorphic sites must also pass a final 'footprint test'. In this test, normalized intensities for probes matching the reference sequence across positions -5 to $+5$ were separately averaged for scans that resulted in reference base calls and

alternate base calls. The normalization step adjusts for systematic differences in brightness between scans. An SNP is rejected if the ratio of mean normalized intensity around reference calls to mean normalized intensity around alternate calls is less than 1.5. The footprint test requires a cumulative analysis of a complete set of arrays of the same design. We required at least one consensus reference call and one alternate call to define a polymorphism; positions with no reference calls were rejected. Once a site is determined to be polymorphic, we used relaxed base calls in other samples to determine if they match the reference or alternate bases. This strategy allows us to maintain a low false-positive rate for SNP detection, while maximizing the amount of usable genotype data obtained for each position identified as a likely polymorphism.

Trace files. Each trace file represents data for one contiguous fragment of tiled sequence in one orientation. Unlike a conventional sequencing trace, there is a one-to-one correspondence between the trace amplitudes and sequence positions. Data for the reverse tiling were reverse complemented before the trace files were generated, so that the forward and reverse reads were both reported for the forward strand of the reference sequence. The trace files contain all the experimental information required to apply our SNP discovery algorithm. However, owing to round-off of the intensity data in the traces, there may be minor differences in results obtained from the trace files compared with using the raw data.

Each trace also carries descriptive information. Supplementary Table 10 explains how to interpret these descriptive fields for Perlegen resequencing traces, and can be used to supplement the NCBI Trace Archive documentation. Using this information, a user can group together all traces from the same scanned image, and all scans from a single hybridization experiment. A limitation of the NCBI Trace Archive is that only a subset of trace file features is searchable through the web interface, and there is no direct way to identify and retrieve traces corresponding to a given genomic interval. However, one can use the Perlegen Sciences Mouse Genome Browser to map submitted trace files to the NCBI Mouse Build 36 reference genomic sequence (Supplementary Fig. 4). The trace files can also be viewed using the Staden package for manipulating trace data (available at <http://staden.sourceforge.net/>). Although we report only high quality SNPs, deposited sequence traces may allow researchers interested in identifying additional polymorphisms to do so.

Quality scores. We developed a method for determining quality scores for each submitted base to make the trace data for individual scans more useful, using an algorithm similar to *Phred*²⁵. For each sequence position, the algorithm uses the intensity ratio between the brightest and next-brightest probes at that position, plus the consistency of the called sequence with the reference sequence across a sliding window around that position. The *Phred* algorithm effectively determines a decision tree that takes these input parameters and has as its output a quality score. Because the decision tree has a limited number of nodes, only a limited number of discrete quality scores are available. Like *Phred* didoxy sequencing quality scores, the reported scores represent an estimate of $-10\log_{10}(P)$ where P is the probability of an incorrect genotype call; hence a score of 20 corresponds to a probability of 0.01, 30 corresponds to 0.001, and so on. The quality scores are not directly used in our SNP discovery algorithm, although the algorithm does use the same underlying features (intensity ratios and local conformance). Owing to experimental variation across individual hybridization experiments, the quality scores for a specific scan in the large-scale study reported here may not be perfectly calibrated, but still provide a measure of relative data quality.

Chromosome and strain average quality scores. To calculate the average quality score by chromosome or strain, the long-range primer pair sequences were mapped to NCBI Build 36 of the mouse sequence. For each amplicon, the average quality score value of the corresponding sequences tiled on the arrays was calculated. Then these quality scores were rolled up by chromosome and/or strain, taking into account the length of each amplicon using the formula: $\sum(\text{amplicon size} \times \text{average quality score for amplicon}) / \sum(\text{amplicon size})$.

The 19 autosomes were all roughly equally well represented, with sequences submitted for 60% of the available bases (Supplementary Table 2). However, for the X and Y chromosome, we only submitted sequences for 40% and 15% of the basepairs respectively. This difference is largely because the sex chromosomes have a much higher proportion of repetitive sequence than the autosomes and the Y reference chromosome presently has large unsequenced gaps. Additionally,

fewer SDP patterns were observed on the X chromosome (12 SDP over 5 Mb) and the Y chromosome (5.5 SDP over 1 Mb), which is concordant with the reduced recombination experienced by these chromosomes, the history of hybrid sterility among the *Mus musculus* subspecies and the observed low SNP-rates for these chromosomes.

We assayed the complete 1.49 billion bases of genomic sequence for all 15 inbred strains (Supplementary Table 11). The average sequence quality scores were similar for the 11 classical inbred strains and the WSB/EiJ wild-derived strain. Owing to the increased sequence divergence of CAST/EiJ, MOLF/EiJ and PWD/PhJ from the C57BL/6J reference strain, there were greater numbers of mismatches between the target DNA and the tiled sequence, resulting in slightly lower average sequence quality scores for these strains.

Generation of ancestral haplotype map. Each pair of classical inbred strains was compared using a hidden Markov model with two states for each SNP in the pair and each SNP position was assigned one of two states corresponding to regions of low SNP-rate (common ancestry) or high SNP-rate (divergent ancestry) intervals using the Viterbi algorithm²⁶. The state transition probabilities between SNPs is $(1 - e^{-\theta L})/2$ where $\theta = 10^{-10}$ and L is the distance between SNPs in bases, which corresponds to a Poisson distribution assuming symmetric transitional probabilities per base of 10^{-10} . In high SNP-rate regions, the probability of observing different SNP alleles in the two strains is $2p(1-p)$ where p is the fraction of strains which have the SNP allele. In low SNP-rate intervals, the probability of observing different SNP alleles is $0.05 \times 2p(1-p)^2$. We only consider intervals longer than 5 kb in the analysis to avoid any bias caused by short intervals. Any SNP where the allele is missing in either strain is labelled as missing in the pair. States of missing SNPs were inferred if the neighbouring states were consistent, which reduced the amount of missing states by 90% leaving only stretches of missing states where a transition between states occurred. A map of all ancestry breakpoints among all strains was created by merging the locations from all pairs of strains where a transition between low and high SNP-density regions occurred. The remaining missing states were inferred by selecting the location of the breakpoint to minimize the total number of breakpoints using a greedy algorithm.

Ancestral origins of haplotypes were predicted using the 5,172,553 SNPs successfully genotyped in all four wild stains. Ancestral origin of each classical inbred strain was assigned using hidden a Markov model with 5 states per SNP position, where one state represents ancestral origin of each wild-derived strain and one state represents unknown ancestral origin. The probability of observing a classical inbred SNP allele matching the SNP allele in the corresponding ancestral strain is 0.9 and the probability of observing any SNP allele in the unknown ancestral state is 0.5. The transition probability to the same state is $(1 + 4e^{-\theta L})/5$ where $\theta = 10^{-10}$ and L is the distance between SNPs in bases. The transition probability to a different state is $(1 - e^{-\theta L})/5$. The forward-backward algorithm assigned a probability to each ancestral state for each SNP. Each SNP where the unknown ancestry probability was higher than 0.5 was assigned unknown ancestry. Each position where the ancestry for a wild-derived strain had probability higher than 0.9 was assigned the corresponding ancestral origin. Remaining positions were assigned ambiguous origin. Ancestry for the MOLF/EiJ strain was predicted by applying a similar approach using a hidden Markov model with four states for each SNP. Three of the states correspond to ancestral origin of the remaining three wild-derived strains and the fourth state represents unknown ancestral origin. We applied the forward-backward algorithm using the same transition and SNP matching probabilities and the same procedure to assign ancestry as above.

Strain distribution patterns. SNPs that were polymorphic among the inbred strains were binned within window size. Window sizes tested were 50-kb, 100-kb, 500-kb, 1-Mb and 5-Mb (Supplementary Table 5). SNPs were included in the analysis if they had a genotype recorded for all 11 classical inbred mouse strains. For each SNP locus, the array of alleles across strains was recorded and a tally was made of the number of SNPs corresponding to the same pattern (array) within windows. Because the number of SDPs can artificially increase owing to genotype errors in the data, we report the number of patterns accounting for 95% of the SNPs for each window size. Genome-wide mean and standard error of SDP count, and SNP count within the different window sizes within and across chromosomes, were calculated (Supplementary Table 5).

LETTERS

The detection of carbonation by the *Drosophila* gustatory system

Walter Fischler^{1,2}, Priscilla Kong^{1,2}, Sunanda Marella^{1,2} & Kristin Scott^{1,2}

There are five known taste modalities in humans: sweet, bitter, sour, salty and umami (the taste of monosodium glutamate). Although the fruitfly *Drosophila melanogaster* tastes sugars, salts and noxious chemicals, the nature and number of taste modalities in this organism is not clear. Previous studies have identified one taste cell population marked by the gustatory receptor gene *Gr5a* that detects sugars, and a second population marked by *Gr66a* that detects bitter compounds^{1–4}. Here we identify a novel taste modality in this insect: the taste of carbonated water. We use a combination of anatomical, calcium imaging and behavioural approaches to identify a population of taste neurons that detects CO₂ and mediates taste acceptance behaviour. The taste of carbonation may allow *Drosophila* to detect and obtain nutrients from growing microorganisms. Whereas CO₂ detection by the olfactory system mediates avoidance⁵, CO₂ detection by the gustatory system mediates acceptance behaviour, demonstrating that the context of CO₂ determines appropriate behaviour. This work opens up the possibility that the taste of carbonation may also exist in other organisms.

To identify uncharacterized taste neurons and examine their selectivity, we undertook an anatomical screen of a Gal4 enhancer trap library (U. Heberlein, unpublished results). This screen identified a Gal4 enhancer trap line, E409, which is expressed in taste peg neurons. The proboscis labellum contains gustatory neurons that innervate taste bristles and taste pegs. The gustatory receptor genes *Gr5a* and *Gr66a* are expressed in neurons underlying taste bristles^{1–3}. *Gr5a* is also expressed in taste peg neurons^{1,2}. E409 marks a taste cell population that does not contain *Gr5a* or *Gr66a* (Fig. 1a).

Taste neurons of the proboscis send axons to the subesophageal ganglion (SOG) of the fly brain (Supplementary Fig. 1)⁶. The *Gr5a* and *Gr66a* neurons show stereotyped, segregated projection patterns, forming a map of taste quality and taste organ^{2,3}. *Gr5a* axons project to the anterior SOG, whereas *Gr66a* neurons project to the medial SOG in a webbed pattern. In contrast, E409 axons terminate in snake-like projections in the anterior SOG, do not cross the midline, and are spatially segregated from *Gr5a* and *Gr66a* projections (Fig. 1b). Interestingly, E409-positive fibres follow a few *Gr5a*-positive axons (presumably from taste pegs) but the terminations are not identical (Fig. 1c). The selective expression in taste cell subsets and the convergence of E409 projections suggest that the E409 enhancer trap line labels a new class of taste neurons.

To determine the ligand specificity of these neurons directly, we expressed the genetically encoded calcium sensor G-CaMP in E409 neurons, stimulated the proboscis with a broad selection of taste ligands and visualized taste-induced activation of E409 projections in the living fly using confocal microscopy⁴. E409 neurons show no activity in response to sugar or bitter compounds, demonstrating that these neurons are both functionally and anatomically distinct (Fig. 2a). In addition, taste cues perceived as different taste modalities

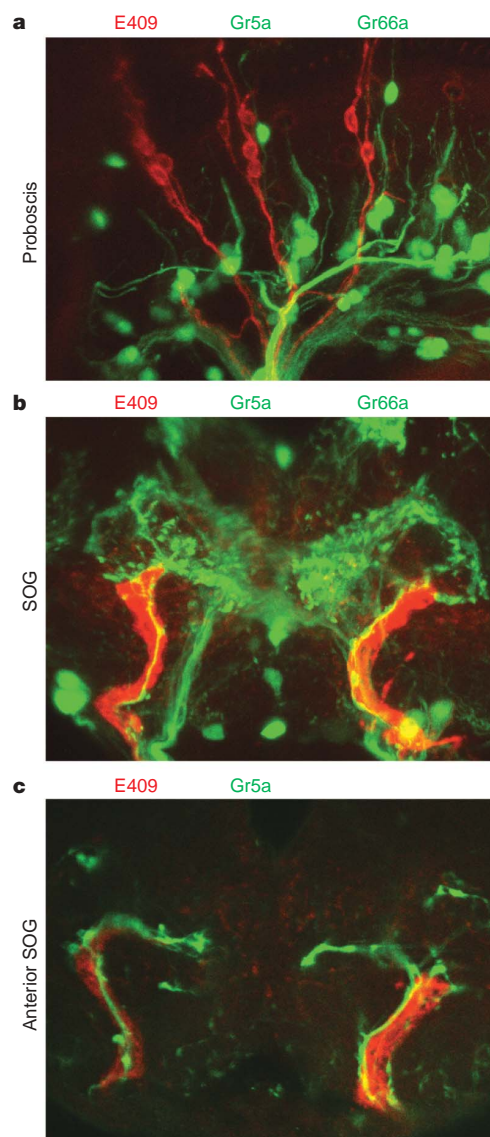


Figure 1 | The E409-Gal4 enhancer trap labels taste peg neurons that project to the taste region of the fly brain. **a**, E409 neurons do not contain markers for sugar or bitter neurons. Shown is a proboscis with E409 neurons (red) and *Gr5a* and *Gr66a* neurons (green). **b**, E409 projections are spatially segregated from other taste projections. Shown is a 60-μm optical section through the SOG with E409 projections (red) and *Gr66a* and *Gr5a* projections (green). **c**, E409 projections (red) terminate differently from *Gr5a* projections (green) in the anterior SOG (10-μm optical section). Scale bar at bottom right, 50 μm.

¹Department of Molecular and Cell Biology, ²Helen Wills Neuroscience Institute, 291 Life Sciences Addition, University of California-Berkeley, Berkeley, California 94720, USA.

reduced preference for soluble CO₂, whereas their genetically identical siblings raised at 22 °C show wild-type preference (Fig. 3a). Wild-type controls raised at 30 °C show normal preference (preference index PI = 0.20 ± 0.09 (22 °C); PI = 0.24 ± 0.07 (30 °C), *t*-test, *P* = 0.7), arguing that rearing conditions do not influence the behaviour (PI is defined in Methods). E409 homozygotes show wild-type preference (PI = 0.22 ± 0.07, *t*-test, *P* = 0.6), arguing that the Gal4 insertion alone does not influence the behaviour. The conditional inactivation experiments suggest that wild-type flies are attracted to CO₂ in solution and that this attraction is mediated by E409 neurons. The preference is weak compared to sucrose preference (100 mM sucrose versus H₂O, E409-Gal4, UAS-Kir2.1, tub-Gal80ts flies, PI = 0.87 ± 0.04 (22 °C), PI = 0.91 ± 0.01 (30 °C)), suggesting that CO₂ may enhance taste preference, but on its own modestly promote taste acceptance. This may be rationalized by considering that CO₂ alone will not provide nutrients and its value may increase in conjunction with food sources (that is, growing yeast).

One limitation with the loss-of-function studies is that the E409-Gal4 line also drives expression in the central nervous system (Supplementary Fig. 1); therefore, decreased preference for CO₂ could be due to central or peripheral defects. To determine whether peripheral E409 neurons are sufficient to induce taste behaviours, we used an inducible activation approach. A modified mammalian ion channel, the VR1E600K capsaicin receptor, was expressed in E409 neurons to enable inducible activation^{4,9}. Previously, we showed that flies expressing VR1E600K in sugar-sensing neurons show acceptance behaviour to capsaicin, whereas expression in bitter-sensing cells mediates avoidance⁴. As expected, E409 neurons expressing the VR1E600K channel exhibit capsaicin-induced G-CaMP fluorescence increases (Fig. 3c). More importantly, flies with the VR1E600K channel in E409 neurons show a marked dose-dependent behavioural preference for capsaicin (Fig. 3d). As capsaicin is accessible only to peripheral neurons, these results establish that activation of E409 sensory neurons is sufficient to trigger taste acceptance behaviour.

This work identifies a taste cell population that is activated by carbon dioxide in water and mediates taste acceptance behaviour. By contrast, CO₂ sensing by olfactory neurons mediates avoidance behaviour in *Drosophila*⁸. Do the taste and smell of CO₂ interact to influence the fly's behaviour? The E409 gustatory neurons are activated by high levels of gas-phase CO₂, and are thus unlikely to influence olfactory avoidance behaviour. To directly test this, we used an olfactory assay to monitor behaviour to 1% gas-phase CO₂ (Fig. 4a). Consistent with previous studies, wild-type flies avoid 1% CO₂ and this avoidance is mediated by the antenna⁵. Flies in which E409 neurons have been genetically silenced show wild-type behaviour, demonstrating that E409 neurons do not participate in olfactory avoidance behaviour (Fig. 4a). To examine whether olfactory neurons influence taste acceptance behaviour, flies with and without antennae were given a choice between sodium bicarbonate solutions at pH 6.5 and at pH 8.5 (Fig. 4b). Flies lacking antennae prefer sodium bicarbonate pH 6.5, demonstrating that olfactory neurons do not influence taste acceptance behaviour. Taken together, these experiments suggest that CO₂ detection by the proboscis and the antennae can act independently to control behaviour.

Recently, the gustatory receptors Gr21a and Gr63a have been identified as the *Drosophila* olfactory CO₂ receptor^{10,11}. These receptors are not detected in E409 neurons^{10,12}, and Gr63a mutants show wild-type preference to soluble CO₂ (Fig. 4b), arguing that the molecular mechanisms underlying olfactory and gustatory CO₂ detection are also distinct. Candidate molecules that may participate in CO₂ taste detection include members of the gustatory receptor gene family^{12–15}, carbonic anhydrases¹⁶, or guanylate cyclases^{17–19}. Identifying the receptor in E409 neurons will provide insight into the diversity of molecular mechanisms that *Drosophila* use to detect CO₂.

The observations that CO₂ in the air mediates avoidance whereas CO₂ in solution mediates acceptance suggest that a single compound can mediate different behaviours, depending on the sensory neurons it activates. CO₂-sensing by the proboscis may mediate detection of

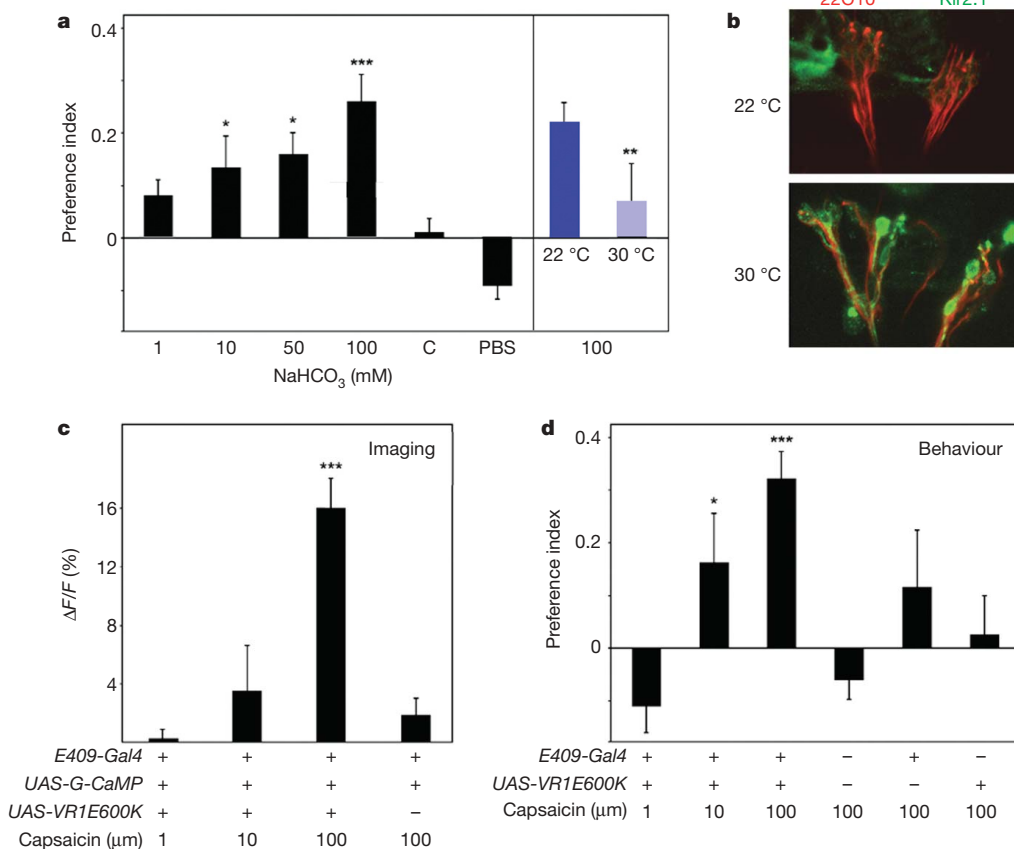


Figure 3 | E409 neurons are necessary for behavioural preference for sodium bicarbonate, pH 6.5, and sufficient to trigger taste acceptance behaviour.

a, Wild-type flies prefer sodium bicarbonate at pH 6.5 versus pH 8.5. Flies show no preference for 100 mM NaHCO₃ pH 6.5 versus pH 6.5 (C). Flies do not prefer PBS pH 6.5 to PBS pH 8.5 (mean + s.e.m. for *n* = 5–6 trials per compound, *t*-test, versus C, **P* < 0.05, ****P* < 0.005). **b**, E409-Gal4, UAS-Kir2.1, tub-Gal80ts flies prefer 100 mM NaHCO₃, pH 6.5, at permissive temperature (22 °C), but not at 30 °C (mean + s.e.m. for *n* = 10 trials per concentration, *t*-test, ***P* < 0.01). **c**, The proboscis of E409-Gal4, UAS-Kir2.1, tub-Gal80ts flies does not express Kir2.1 (green) at 22 °C but does at 30 °C. Counter-stain 22C10 (red), scale 57 × 82 μm. **d**, Flies containing VR1E600K in E409 neurons show capsaicin-induced G-CaMP fluorescence increases (mean + s.e.m. for *n* = 5 flies per concentration). **e**, Flies with VR1E600K in E409 neurons prefer capsaicin (mean + s.e.m. for *n* = 5 trials per concentration; *t*-test, versus *w*1118 100 μM capsaicin, **P* < 0.05, ****P* < 0.005).

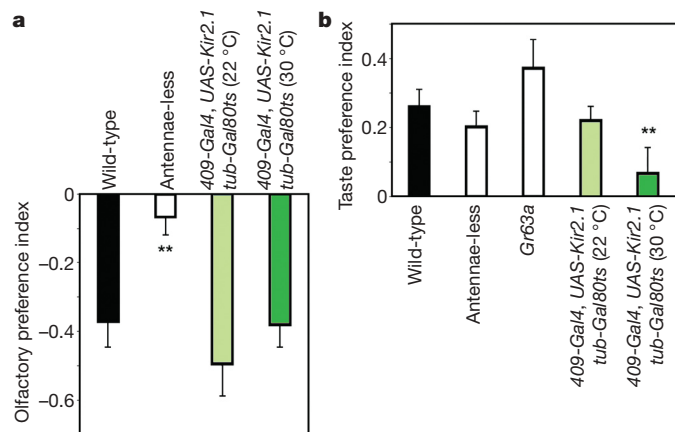


Figure 4 | Segregation of CO₂ detection by the taste and olfactory systems. **a**, E409 neurons do not participate in CO₂ olfactory avoidance behaviour. Wild-type and *E409-Gal4, UAS-Kir2.1, tub-Gal80ts* flies at permissive and restrictive temperature avoid 1% gas-phase CO₂. Flies lacking antennae do not avoid CO₂ (mean + s.e.m. for $n = 4$ –10 trials per genotype, t -test, $**P < 0.01$). **b**, Olfactory neurons do not participate in CO₂ taste behaviour. Flies with and without antennae, *Gr63a* mutants, and *E409-Gal4, UAS-Kir2.1, tub-Gal80ts* flies at permissive temperature prefer NaHCO₃, pH 6.5 versus pH 8.5, whereas *E409-Gal4, UAS-Kir2.1, tub-Gal80ts* flies at 30 °C are neutral (mean + s.e.m. for $n = 5$ –10 trials per genotype, t -test, $**P < 0.01$). The Kir2.1 data from Fig. 3b are shown for reference.

microorganisms such as growing yeast, allowing fruitflies to identify and consume nutrients that these microorganisms provide. CO₂-sensing by the antenna may signify unfavourable conditions such as hypoxia, macroorganisms, overly rotten fruit or stressful environments^{5,20}. Thus, compartmentalization of CO₂ detection may allow fruitflies to distinguish local versus global CO₂ levels and finely regulate behaviour.

Finally, our finding that E409 neurons respond to carbonated water and not to one of the classic taste modalities argues that carbon dioxide (or carbonic acid) is a novel taste category in *Drosophila*. An intriguing possibility is that CO₂ is a species-specific taste, such that organisms possess taste neurons tuned to chemicals over-represented in the local environment or necessary for survival. Thus, taste modalities may differ according to nutritional needs. Alternatively, CO₂ may be an unappreciated taste modality in many organisms.

METHODS SUMMARY

Transgenic flies and immunohistochemistry. *Gr5a-GFP-IRES-GFP-IRES-GFP* transgenic flies were generated using standard procedures^{3,6}. Double labelling of the proboscis and brain was performed as previously described³.

G-CaMP imaging experiments. Imaging studies were performed as described⁴. Taste compounds and concentrations are given in Methods. Five flies per compound were tested for a mean response (+s.e.m.). For the concentration curve, 0.5 ml of 0.1 mM, 1 mM, 10 mM, 50 mM, 100 mM and 200 mM sodium bicarbonate in H₂O were prepared to which 50 μ l of 5 M H₂PO₄ was added. Caesium carbonate was buffered to pH 7 with HCl.

Behavioural assays. An existing taste choice assay was modified to measure the behaviour of flies to solutions⁴. The number of flies on each substrate was recorded and counted at 3 min intervals. The taste preference index was calculated as PI = (number of flies on test substance – number of flies on control substance)/(total number of flies on the substrates). Five–ten batches of 100–200 flies were tested to yield a mean PI (\pm s.e.m.). The olfactory preference index was calculated as PI = (number of flies on 1% CO₂ injected side – number of flies on air-injected side)/(total number of flies). Each test consisted of 50–60 flies, 4–10 batches per genotype. For *E409-Gal4, UAS-Kir2.1, tub-Gal80ts* experiments, newly eclosed flies were either grown at 18 °C for 2 days and starved overnight at room temperature, or grown and starved overnight at 30 °C. Behavioural assays were performed at 22 °C.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 18 June; accepted 17 July 2007.

- Chyb, S., Dahanukar, A., Wickens, A. & Carlson, J. R. *Drosophila* Gr5a encodes a taste receptor tuned to trehalose. *Proc. Natl Acad. Sci. USA* **100** (suppl. 2), 14526–14530 (2003).
- Thorne, N., Chromey, C., Bray, S. & Amrein, H. Taste perception and coding in *Drosophila*. *Curr. Biol.* **14**, 1065–1079 (2004).
- Wang, Z., Singhvi, A., Kong, P. & Scott, K. Taste representations in the *Drosophila* brain. *Cell* **117**, 981–991 (2004).
- Marella, S. et al. Imaging taste responses in the fly brain reveals a functional map of taste category and behavior. *Neuron* **49**, 285–295 (2006).
- Suh, G. S. et al. A single population of olfactory sensory neurons mediates an innate avoidance behaviour in *Drosophila*. *Nature* **431**, 854–859 (2004).
- Stocker, R. F. The organization of the chemosensory system in *Drosophila melanogaster*: a review. *Cell Tissue Res.* **275**, 3–26 (1994).
- Baines, R. A., Uhler, J. P., Thompson, A., Sweeney, S. T. & Bate, M. Altered electrical properties in *Drosophila* neurons developing without synaptic transmission. *J. Neurosci.* **21**, 1523–1531 (2001).
- McGuire, S. E., Mao, Z. & Davis, R. L. Spatiotemporal gene expression targeting with the TARGET and gene-switch systems in *Drosophila*. *Sci. STKE* **2004**, pl6 (2004).
- Jordt, S. E., Tominaga, M. & Julius, D. Acid potentiation of the capsaicin receptor determined by a key extracellular site. *Proc. Natl Acad. Sci. USA* **97**, 8134–8139 (2000).
- Jones, W. D., Cayirlioglu, P., Kadow, I. G. & Vosshall, L. B. Two chemosensory receptors together mediate carbon dioxide detection in *Drosophila*. *Nature* **445**, 86–90 (2007).
- Kwon, J. Y., Dahanukar, A., Weiss, L. A. & Carlson, J. R. The molecular basis of CO₂ reception in *Drosophila*. *Proc. Natl Acad. Sci. USA* **104**, 3574–3578 (2007).
- Scott, K. et al. A chemosensory gene family encoding candidate gustatory and olfactory receptors in *Drosophila*. *Cell* **104**, 661–673 (2001).
- Clyne, P. J., Warr, C. G. & Carlson, J. R. Candidate taste receptors in *Drosophila*. *Science* **287**, 1830–1834 (2000).
- Dunipace, L., Meister, S., McNealy, C. & Amrein, H. Spatially restricted expression of candidate taste receptors in the *Drosophila* gustatory system. *Curr. Biol.* **11**, 822–835 (2001).
- Robertson, H. M., Warr, C. G. & Carlson, J. R. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **100** (Suppl. 2), 14537–14542 (2003).
- Coates, E. L. Olfactory CO₂ chemoreceptors. *Respir. Physiol.* **129**, 219–229 (2001).
- Gray, J. M. et al. Oxygen sensation and social feeding mediated by a *C. elegans* guanylate cyclase homologue. *Nature* **430**, 317–322 (2004).
- Verma, A., Hirsch, D. J., Glatt, C. E., Ronnett, G. V. & Snyder, S. H. Carbon monoxide: a putative neural messenger. *Science* **259**, 381–384 (1993).
- Wingrove, J. A. & O'Farrell, P. H. Nitric oxide contributes to behavioral, cellular, and developmental responses to low oxygen in *Drosophila*. *Cell* **98**, 105–114 (1999).
- Faucher, C., Forstreuter, M., Hilker, M. & de Bruyne, M. Behavioral responses of *Drosophila* to biogenic levels of carbon dioxide depend on life-stage, sex and olfactory context. *J. Exp. Biol.* **209**, 2739–2748 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank U. Heberlein and her laboratory for generating and providing the Gal4 enhancer trap library containing the *E409-Gal4* transgenic flies; Z. Wang for generation of the *Gr5a-GFP-IRES-GFP-IRES-GFP* transgenic flies; S. Asgarian for technical assistance in the anatomy screen; and G. Agarwal for assistance with Matlab. We also thank L. Vosshall, C. Zuker and members of the Scott laboratory for providing comments on the manuscript. This work was supported by a grant from the NIH (NIDCD), a Burroughs Wellcome Fund Career Award, a McKnight Scholar Award and a John Merck Award to K.S.

Author Contributions W.F. performed the majority of G-CaMP imaging experiments, developed the behavioural assay, performed behavioural experiments and co-wrote the manuscript. P.K. performed the anatomy screen leading to the identification of *E409-Gal4*. S.M. performed the G-CaMP imaging experiments of capsaicin-induced responses in *E409-Gal4, UAS-GCaMP, UAS-VRIE600K* flies, sequenced the Gal4 insertion site in the *E409-Gal4* flies and participated in initial characterization of the E409 neurons. K.S. assisted P.K. in the anatomy screen and W.F. in the G-CaMP imaging and behavioural studies, co-wrote the manuscript and supervised the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.S. (kscott@berkeley.edu).

METHODS

Transgenic flies. *Gr5a-GFP-IRES-GFP-IRES-GFP* transgenic flies were generated by cloning an 8.5-kb fragment upstream of *Gr5a* into a transformation vector containing GFP-IRES-GFP-IRES-GFP (ref. 3), and transformant flies were generated using standard procedures.

The *E409-Gal4* insertion sequence is AE003525, position 26026, upstream of the limpet gene (CG32171, FBgn0036672; <http://flybase.bio.indiana.edu/reports/FBgn0036672.html>), encoding a LIM homeo-domain protein. The limpet gene is not expressed in taste neurons by *in situ* hybridization (data not shown). Another enhancer trap line, NP1017, is expressed in taste peg neurons and labellar neurons, and was recently reported to mark water-sensing cells in *Drosophila*²¹. These studies examined the response of labellar neurons only and did not examine response profiles of taste peg neurons. We examined the responses of taste peg neurons labelled by NP1017 and find that they respond to carbonated water but not to water (data not shown), suggesting that NP1017 and *E409* label the same population of taste peg neurons. *UAS-Kir2.1* contains *Kir2.1* tethered to GFP (ref. 8).

Immunohistochemistry. Double labelling of the proboscis and brain was performed as previously described³. In Fig. 1, CD2 (red) and GFP (green) reporters were detected by immunohistochemistry on flies containing *E409-Gal4*, *UAS-CD2*, *Gr66a-GFP-IRES-GFP-IRES-GFP*, *Gr5a-GFP-IRES-GFP-IRES-GFP* transgenes. In Fig. 3b, *Kir2.1* expression is detected by GFP immunohistochemistry of *Kir2.1-GFP* (green). Taste peg neurons are counter-stained with 22C10 antibody 1:20 (red), which labels the microtubule-associated protein Futsch²².

G-CaMP imaging experiments. Imaging studies were performed as described⁴. Taste concentrations used were 1 M arabinose, fructose, galactose, glucose, maltose, sucrose, trehalose, NaCl, KCl, ascorbic acid, citric acid; 100 mM caffeine, HCl, alanine, arginine, aspartic acid, cysteine, monosodium glutamate, glutamic acid, histidine, leucine, lysine, methionine, serine, threonine, valine; 10 mM aristolochic acid, berberine, denatonium, lobeline, papaverine, quinine, HCl, NaCl; 1 mM HCl; 0.7 mM azadirachtin; 0.1 mM limonin; 5% yeast plus 10 mM sucrose; 1% acetic acid, formic acid, EtOH; 0.1 mg ml⁻¹ quassin. Data were collected from flies that responded to carbonated water or yeast at the end of the experiment. Five flies per compound were tested for a mean response (\pm s.e.m.). To determine statistical significance, responses were compared to the water response.

For the concentration curve, solutions of 0.1 mM, 1 mM, 10 mM, 50 mM, 100 mM and 200 mM sodium bicarbonate in H₂O were prepared to which H₂PO₄ buffer (500 mM) was added to lower the pH of the solutions (pH = 4.5–6.5). For each concentration, 50 μ l of 5 M H₂PO₄ was added to 0.5 ml of sodium bicarbonate. This led to slightly different pH values and percentage CO₂ conversion at different concentrations (0.1 mM = pH 4.5, 99% CO₂ species; 1 mM = pH 5, 96% CO₂ species; 10 mM = pH 5, 96% CO₂ species; 50 mM =

pH 5.5, 88% CO₂ species; 100 mM = pH 6, 70% CO₂ species; and 200 mM = pH 6.5, 42% CO₂ species). Control solutions of 200 mM sodium bicarbonate in H₂O exist at a pH of 9 at which 0.2% of species are CO₂. Caesium carbonate was buffered to pH 7 with HCl. Carbon dioxide levels were measured and calculated using an Orion carbon dioxide electrode and manual (Thermal Electron Corporation).

Behavioural assays. We modified an existing taste choice assay to measure the behaviour of flies in response to solutions⁴. Briefly, ~100–200 flies were placed in a closed box in which the top surface contains 4 cottonballs in empty vials. The cotton was soaked in different solutions (8 ml per cottonball), two with the test compound and two with a control solution. The number of flies on each substrate was recorded for a 60 min period using the Logitech Image Studio digital camera and software, and counted at 3 min intervals. The taste preference index was calculated as PI = (number of flies on test substance – number of flies on control substance)/(total number of flies on the substrates). Five–ten batches of 100–200 flies were tested to yield a mean PI (\pm s.e.m.). For the *E409-Gal4*, *UAS-Kir2.1*, *tub-Gal80ts* experiments, newly eclosed flies were either grown at 18 °C for 2 days and starved overnight at room temperature, or grown and starved overnight at 30 °C to inactivate the *E409* neurons. For all other fly lines, flies were grown at room temperature for 3–5 days and starved overnight. All behavioural assays were performed at 22 °C.

For the olfactory assay, flies were placed on ice for 1 min and then put in a circular plastic chamber (constructed by P. Masek) with a diameter of 50 mm and height of 2 mm. The chamber has a perforated floor with a lower chamber into which two airstreams (air versus 1% CO₂ in air) are injected from opposite sides. A dividing wall prevents mixing of air from the two airstreams in the lower chamber. The number of flies on each side of the chamber was recorded for a 10 min period using a Logitech camera and software and counted at 30 s intervals. The olfactory preference index was calculated as PI = (number of flies on 1% CO₂ injected side – number of flies on air-injected side)/(total number of flies). Each test consisted of 50–60 flies. Four batches of the *w¹¹¹⁸* and *w¹¹¹⁸*-antennae flies and ten batches of the *E409-Gal4*, *UAS-Kir2.1*, *tub-Gal80ts* flies were tested to yield a mean PI. As above, for the *E409-Gal4*, *UAS-Kir2.1*, *tub-Gal80ts* experiments, newly eclosed flies were either grown at 18 °C for 2 days and starved overnight at room temperature, or grown and starved overnight at 30 °C to inactivate the *E409* neurons. For all other fly lines, flies were grown at room temperature for 3–5 days and starved overnight.

21. Inoshita, T. & Tanimura, T. Cellular identification of water gustatory receptor neurons and their central projection pattern in *Drosophila*. *Proc. Natl Acad. Sci. USA* **103**, 1094–1099 (2006).
22. Hummel, T., Kruckert, K., Roos, J., Davis, G. & Klammt, C. *Drosophila* Futsch/22C10 is a MAP1B-like protein required for dendritic and axonal development. *Neuron* **26**, 357–370 (2000).

LETTERS

Dominant-negative mutations in the DNA-binding domain of STAT3 cause hyper-IgE syndrome

Yoshiyuki Minegishi¹, Masako Saito¹, Shigeru Tsuchiya², Ikuya Tsuge³, Hidetoshi Takada⁴, Toshiro Hara⁴, Nobuaki Kawamura⁵, Tadashi Ariga⁵, Srdjan Pasic⁶, Oliver Stojkovic⁷, Ayse Metin⁸ & Hajime Karasuyama¹

Hyper-immunoglobulin E syndrome (HIES) is a compound primary immunodeficiency characterized by a highly elevated serum IgE, recurrent staphylococcal skin abscesses and cyst-forming pneumonia, with disproportionately milder inflammatory responses, referred to as cold abscesses, and skeletal abnormalities¹. Although some cases of familial HIES with autosomal dominant or recessive inheritance have been reported, most cases of HIES are sporadic, and their pathogenesis has remained mysterious for a long time. Here we show that dominant-negative mutations in the human signal transducer and activator of transcription 3 (STAT3) gene result in the classical multisystem HIES. We found that eight out of fifteen unrelated non-familial HIES patients had heterozygous STAT3 mutations, but their parents and siblings did not have the mutant STAT3 alleles, suggesting that these were *de novo* mutations. Five different mutations were found, all of which were located in the STAT3 DNA-binding domain. The patients' peripheral blood cells showed defective responses to cytokines, including interleukin (IL)-6 and IL-10, and the DNA-binding ability of STAT3 in these cells was greatly diminished. All five mutants were non-functional by themselves and showed dominant-negative effects when co-expressed with wild-type STAT3. These results highlight the multiple roles played by STAT3 in humans, and underline the critical involvement of multiple cytokine pathways in the pathogenesis of HIES.

Elevated serum IgE is a hallmark of many allergic disorders². Curiously enough, the hyper-IgE state is also observed in some primary immunodeficiency disorders, such as HIES, Wiskott-Aldrich syndrome, Omenn syndrome and Comèl-Netherton syndrome³. HIES (OMIM number 243700) was first reported in 1966 as Job's syndrome (OMIM number 147060)^{4,5}, but its underlying cause is still unknown, unlike the other three syndromes. In most cases of HIES, the clinical manifestations extend over multiple systems in the body, including the immune system, skeletal/dental system and soft tissue⁶. In contrast, the abnormalities in familial autosomal recessive (AR)-HIES patients seem to be confined to the immune system⁷. We previously identified a homozygous mutation of the tyrosine kinase 2 (TYK2) gene in a patient who showed AR-HIES and susceptibility to intracellular bacterial infections⁸. TYK2 is a non-receptor tyrosine kinase belonging to the JAK family^{9,10}. The patient's cells expressed no detectable TYK2 protein and displayed defects in multiple cytokine signals, including the signalling pathways for IL-6, IL-10, IL-12, IL-23 and type I IFN. The cytokine signals were successfully restored by introducing the intact TYK2 gene into the patient's cells. These multiple defects probably account for the patient's complex clinical manifestations⁸. The identification of a TYK2 deficiency in this HIES

patient indicated to us that, besides TYK2, one or more molecules shared by multiple cytokine signalling pathways might also cause HIES.

To explore this possibility, we first examined the responses to IL-6, IL-10, IL-12 and IFN α of peripheral blood cells from two patients (patient 1 and patient 2) who showed characteristics of multisystem HIES, including skeletal/dental abnormalities, skin abscesses, cyst-forming pneumonia and highly elevated serum IgE (Supplementary Table 1). The patients' B cells secreted IgM normally when stimulated with Epstein-Barr virus (EBV) infection (Fig. 1a). However, additional stimulation with IL-6 induced no significant increase in IgM secretion in the patients' B cells, unlike in the control B cells (Fig. 1a). Moreover, the suppression of lipopolysaccharide-induced production of TNF α by IL-10 deteriorated in the patients' macrophages (Fig. 1b). Thus, both the IL-6 and IL-10 pathways were defective in these HIES patients, as in the TYK2-deficient patient (Fig. 1a and 1b). In contrast, neither IL-12 nor IFN α signalling was impaired in the HIES patients, unlike in the TYK2-deficient patient (Fig. 1c and 1d). The HIES patients' T cells produced IFN γ normally in response to IL-12 (Fig. 1c), and their peripheral blood mononuclear cells (PBMCs) showed normal upregulation of transcripts for two IFN-inducible genes, *NMI* and *MX1* (also known as *MxA*), in response to IFN α (Fig. 1d). These observations indicated a possible abnormality in one or more molecules that was shared by the IL-6 and IL-10 signals but not essential for the IL-12 and IFN α pathways.

An array of cytokine signals is transduced by different combinations of JAK family kinases and STATs^{9,10}. When cytokines bind to their receptors, receptor-associated JAKs are activated to phosphorylate STATs, which in turn dimerize and translocate to the nucleus, where they activate target genes. In a survey of possible abnormalities in this signal cascade, we identified heterozygous mutations in the STAT3 DNA-binding domain in both patients: a single amino acid deletion (Δ V463) in patient 1 and a mis-sense mutation (R382W) in patient 2 (Fig. 2). Because STAT3 is shown to be activated in response to a wide variety of cytokines, growth factors, and hormones^{11,12}, we thought that its mutation could well account for the patients' complex clinical manifestations extending over multiple systems. Further analysis of the STAT3 complementary DNA sequences in thirteen more unrelated patients with non-familial HIES identified heterozygous mutations in six of these patients (patients 3–8; see Supplementary Table 1 for clinical summary), and all of these mutations were located in the DNA-binding domain of STAT3: Δ V463 in patient 3 and patient 8, like in patient 1; R382Q in patient 4; H437Y in patient 5; T389I in patient 6; R382W in patient 7, like in patient 2 (Fig. 2). The five different STAT3 mutations found in the

¹Department of Immune Regulation, Tokyo Medical and Dental University Graduate School, Tokyo 113-8519, Japan. ²Department of Pediatrics, Tohoku University Graduate School of Medicine, Sendai 980-8575, Japan. ³Department of Pediatrics, Fujita Health University, Aichi 470-1192, Japan. ⁴Department of Pediatrics, Kyushu University, Fukuoka 812-8582, Japan. ⁵Department of Pediatrics, Hokkaido University Graduate School of Medicine, Sapporo 060-8638, Japan. ⁶Pediatric Immunology, Mother and Child Health Institute, Belgrade 110 70, Serbia. ⁷Laboratory for Forensic Genetics, Institute of Forensic Medicine, University of Belgrade, Belgrade 110 70, Serbia. ⁸Pediatric Immunology Department, SB Ankara Diskapi Children's Hospital, Ankara 06110, Turkey.

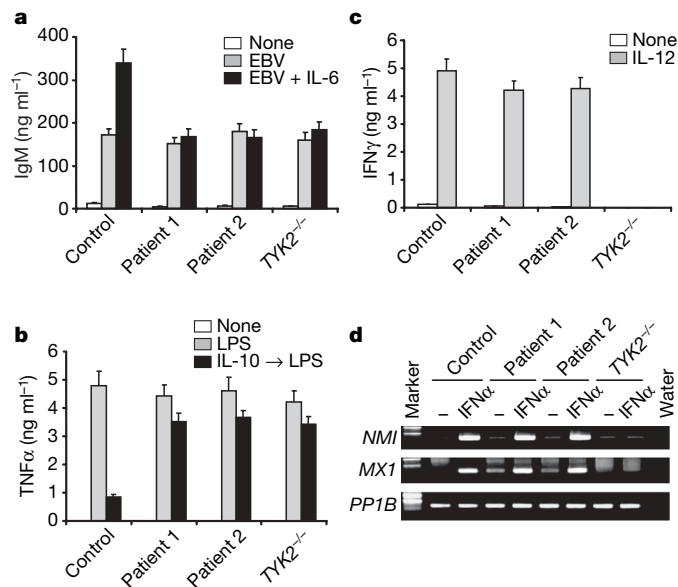


Figure 1 | Impaired responses to IL-6 and IL-10 in HIES patients' cells.

a, IgM levels in culture medium of PBMCs from a control subject, two HIES patients (patient 1 and patient 2), and the TYK2-deficient patient, cultured for 7 days without stimulation, with EBV alone or with EBV and IL-6. **b**, TNFα levels in culture medium of macrophages from the same subjects, cultured without or with lipopolysaccharide (LPS) stimulation for 48 h, or with IL-10 treatment for 24 h before lipopolysaccharide stimulation. **c**, IFNγ in culture medium of CD4⁺ T cells from the same subjects, cultured for 24 h without or with IL-12. Error bars show standard deviations (**a–c**). **d**, NMI, MX1 and cyclophilin B (PP1B) transcript levels in PBMCs from the same subjects were unstimulated or stimulated with IFNα for 2 h.

eight patients were all confirmed by sequencing their genomic DNA, and no sequence alterations were detected in any other parts of *STAT3*, *TYK2* or *JAK1*. The DNA-binding domain of *STAT3* is highly conserved among different species in its amino acid sequence, and the alterations identified in the patients' *STAT3* gene were statistically highly significant in population genetics ($P < 10^{-15}$, Fisher's exact probability test), as judged by the fact that such alterations were not found in 1,000 unrelated healthy individuals analysed, including at least 100 controls from the patients' ethnic group. *STAT3* is located

on human chromosome 17q21, but not 4q, which was reported to contain a disease locus for familial AD-HIES¹³. None of the eight HIES patients in the present study had a known family history of HIES, and no mutation was detected in the *STAT3* cDNAs from all the parents and seven siblings of the patients, even though an analysis of multiple polymorphic markers confirmed the biological parent–child relationship (data not shown). Therefore, the mutations are likely to have occurred *de novo* in the HIES patients.

We next evaluated the biological significance of the *STAT3* mutations. The *STAT3* protein levels were comparable in all the EBV-transformed B-cell lines established from patients 1–6 and a control subject, and the extent of the tyrosine phosphorylation on *STAT3* induced by IFNα stimulation was also comparable (Fig. 3a). Furthermore, all the mutant proteins formed a complex with wild-type *STAT3* as efficiently as did wild-type *STAT3*, as judged by the co-immunoprecipitation of wild-type and mutant *STAT3* proteins co-expressed in COS7 cells (Fig. 3b). However, nuclear extracts isolated from the patients' cells stimulated with IFNα contained much lower amounts of active *STAT3* that could bind to target DNA compared with nuclear extracts from the control cells, whereas the DNA-binding activity of *STAT1* was intact in the patients' cells (Fig. 3c). This finding was consistent with a previous report showing that changing codons 461 to 463 in *STAT3* from Val-Val-Val to Ala-Ala-Ala resulted in the impairment of *STAT3*'s DNA-binding activity¹⁴. Thus, the genetic mutations identified in the HIES patients seemed to result in the impairment of the DNA-binding activity of *STAT3* and most likely that of heterodimers between mutant and wild-type *STAT3* molecules. A similar impairment of DNA-binding activity in *STAT1* protein was recently reported in cells isolated from patients carrying a heterozygous mutation in the DNA-binding domain of *STAT1* (ref. 15).

When wild-type *STAT3* was exogenously expressed—together with a luciferase reporter gene containing *STAT3*-responsive elements—in human HeLa cells in which the endogenous *STAT3* was knocked down, fivefold upregulation of luciferase activity in response to IFNα was detected (Fig. 4a). In contrast, none of the *STAT3* mutants conferred any significant increase of luciferase activity on the HeLa cells in response to IFNα, demonstrating a loss of function of the *STAT3* mutants. To explore the possibility that the *STAT3* mutant proteins function as dominant-negative, wild-type *STAT3* or the individual mutants were exogenously expressed in IL-6-responsive HepG2 cells and IL-10-responsive MC/9 cells (Fig.

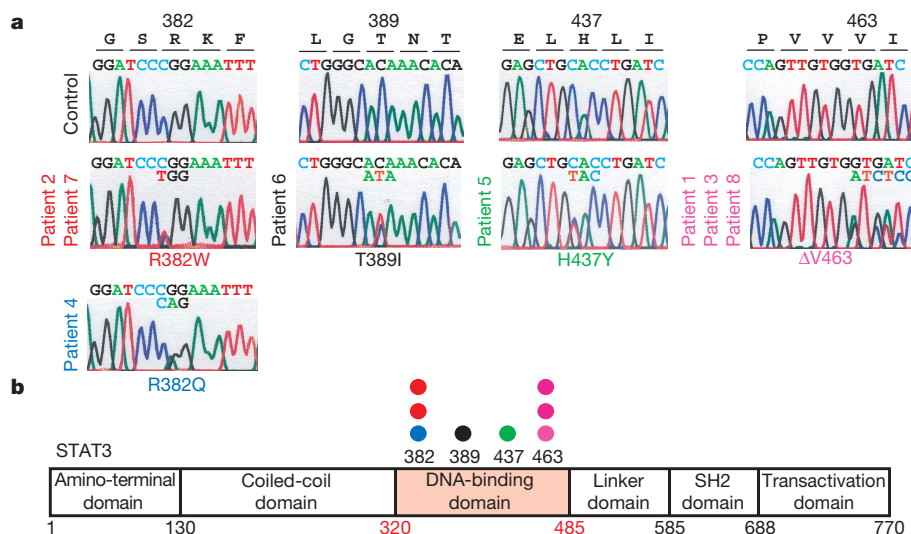


Figure 2 | Heterozygous mutations in the DNA-binding domain of *STAT3* from eight HIES patients. **a**, Electropherograms showing partial *STAT3* cDNA sequences from a control subject and the eight HIES patients. **b**, The

structure of *STAT3* is shown schematically, and the positions of the mutations identified in the eight patients are indicated.

observed in HIES patients. Another remarkable feature of HIES is that patients are often afebrile and feel well¹, despite serious pneumonia or dermal pathology⁴. Indeed, acute-phase responses, such as an increase in serum C-reactive protein during severe infections, were diminished in our patients. STAT3 was originally identified as a protein binding to the IL-6-responsive element in the genes encoding hepatic acute-phase proteins^{21,22}, and the liver-specific inactivation of STAT3 leads to an impaired acute-phase response in mice²³. Thus, the apparent lack of classical inflammatory responses in HIES patients could be attributed to defective signalling of pro-inflammatory cytokines, including IL-6.

Enhanced IgE production in the patients may reflect dysregulated immune responses owing to the impaired response to IL-10, a critical negative regulator²⁴, even though the exact mechanism of hyper IgE remains to be determined, as in the case of other disorders such as Wiskott-Aldrich syndrome. HIES patients often suffer from severe staphylococcal infection in the skin and lung. STAT3 plays a critical part in T_H17 development²⁵, and IL-17 produced by T_H17 cells is protective in the host defence against extracellular bacteria²⁶. IL-22 stimulates cells in the skin and respiratory systems to produce β -defensins through STAT3 activation²⁷. Thus, the susceptibility to bacterial infection could be attributed, at least in part, to the defects in T_H17 development and IL-22 signalling. Among the 15 sporadic HIES patients investigated in this study, no apparent difference was observed in clinical phenotypes and severity between those with the STAT3 mutations and those without the mutations, indicating that other HIES aetiology might be functionally linked to STAT3.

In summary, the present study identified a human deficiency in STAT3 as a major cause of multisystem HIES. This study highlights the multiple and critical roles of STAT3 in humans. The identification of these STAT3 mutations as causative for HIES, in addition to the previous finding of a causative mutation in TYK2 (ref. 8), underlines the critical involvement of a variety of cytokine signals in the pathogenesis of HIES. The diagnosis of HIES early in life is often hampered by a paucity of specific clinical features. Our discovery of STAT3 as a major causative gene of this disease will facilitate earlier and definitive diagnosis, leading to the prevention of serious complications by prompting the start of prophylactic antibiotic treatment early in life.

METHODS SUMMARY

Patients. This study was approved by the institutional review board at the Tokyo Medical and Dental University; written informed consent was obtained from all the individuals studied. The clinical characteristics of the eight HIES patients investigated in this study are summarized in Supplementary Table 1, and all the patients display definitive phenotypes of multisystem HIES (score ≥ 40).

Stimulation of cells with cytokines, and measurement of cytokines and IgM production. Cells were stimulated for the indicated time in culture with cytokines as described previously⁸. The concentration of IFN γ and TNF α in the culture supernatants was determined by ELISA (BD-PharMingen), according to the manufacturer's instructions. The amount of IgM secretion from EBV-infected B cells was determined as previously described²⁸.

RT-PCR and direct sequence analysis. Extraction of total RNA, cDNA synthesis, PCR, semiquantitative RT-PCR, and sequencing were performed as previously described²⁹.

Immunoblotting and immunoprecipitation. Immunoblotting and immunoprecipitation were performed as described previously⁸.

Enzyme-linked DNA-protein interaction assay. Binding of STAT3 and STAT1 to their target DNA was measured using the Mercury TransFactor kit (Clontech Laboratories) according to the manufacturer's protocol.

Retroviral infections. Retroviral infections were done as described previously⁸.

Flow cytometric analysis. The surface immunophenotype was analysed as described³⁰.

STAT3 knock-down. Transfection of short interfering RNA (siRNA) oligonucleotides (5'-ccugcaagagucgauguucucuu-3' and 5'-gcaguuuucagagcag-guauuu-3') was performed as described previously⁸. Forty hours after transfection, the cells were treated with IFN α for 5 h. In the experiment shown

in Fig. 4a, nucleotide sequences of wild-type and mutant STAT3 cDNAs were modified so that they were insensitive to STAT3 siRNA, but they still encoded the original amino acid sequences of STAT3.

Luciferase reporter assay. The reporter construct of STAT3-responsive elements linked to a luciferase reporter gene was transfected with wild-type or mutant STAT3. Forty hours after the transfection, the cells were stimulated with 100 ng ml⁻¹ IL-6 for 5 h. Luciferase activity was measured with a dual-luciferase assay system according to the manufacturer's protocol (Promega).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 June; accepted 19 July 2007.

Published online 5 August 2007; corrected 30 August 2007 (details online).

- Grimbacher, B., Holland, S. M. & Puck, J. M. Hyper-IgE syndromes. *Immunol. Rev.* **203**, 244–250 (2005).
- Gould, H. J. *et al.* The biology of IGE and the basis of allergic disease. *Annu. Rev. Immunol.* **21**, 579–628 (2003).
- Grimbacher, B., Belohradsky, B. H. & Holland, S. M. Immunoglobulin E in primary immunodeficiency diseases. *Allergy* **57**, 995–1007 (2002).
- Davis, S. D., Schaller, J. & Wedgwood, R. J. Job's Syndrome. Recurrent, "cold", staphylococcal abscesses. *Lancet* **1**, 1013–1015 (1966).
- Buckley, R. H., Wray, B. B. & Belmaker, E. Z. Extreme hyperimmunoglobulinemia E and undue susceptibility to infection. *Pediatrics* **49**, 59–70 (1972).
- Grimbacher, B. *et al.* Hyper-IgE syndrome with recurrent infections—an autosomal dominant multisystem disorder. *N. Engl. J. Med.* **340**, 692–702 (1999).
- Renner, E. D. *et al.* Autosomal recessive hyperimmunoglobulin E syndrome: a distinct disease entity. *J. Pediatr.* **144**, 93–99 (2004).
- Minegishi, Y. *et al.* Human tyrosine kinase 2 deficiency reveals its requisite roles in multiple cytokine signals involved in innate and acquired immunity. *Immunity* **25**, 745–755 (2006).
- Schindler, C. & Darnell, J. E. Jr. Transcriptional responses to polypeptide ligands: the JAK-STAT pathway. *Annu. Rev. Biochem.* **64**, 621–651 (1995).
- Ihle, J. N. Cytokine receptor signalling. *Nature* **377**, 591–594 (1995).
- Levy, D. E. & Darnell, J. E. Jr. STATs: transcriptional control and biological impact. *Nature Rev. Mol. Cell Biol.* **3**, 651–662 (2002).
- Kisseleva, T., Bhattacharya, S., Braunstein, J. & Schindler, C. W. Signaling through the JAK/STAT pathway, recent advances and future challenges. *Gene* **285**, 1–24 (2002).
- Grimbacher, B. *et al.* Genetic linkage of hyper-IgE syndrome to chromosome 4. *Am. J. Hum. Genet.* **65**, 735–744 (1999).
- Horvath, C. M., Wen, Z. & Darnell, J. E. Jr. A STAT protein domain that determines DNA sequence recognition suggests a novel DNA-binding domain. *Genes Dev.* **9**, 984–994 (1995).
- Chaplier, A. *et al.* Novel STAT1 alleles in otherwise healthy patients with mycobacterial disease. *PLoS Genet.* **2**, e131 (2006).
- Takeda, K. *et al.* Targeted disruption of the mouse Stat3 gene leads to early embryonic lethality. *Proc. Natl Acad. Sci. USA* **94**, 3801–3804 (1997).
- Darnell, J. E. Jr. STATs and gene regulation. *Science* **277**, 1630–1635 (1997).
- Levy, D. E. & Lee, C. K. What does Stat3 do? *J. Clin. Invest.* **109**, 1143–1148 (2002).
- O'Brien, C. A., Gubrij, I., Lin, S. C., Saylor, R. L. & Manolagas, S. C. STAT3 activation in stromal/osteoblastic cells is required for induction of the receptor activator of NF- κ B ligand and stimulation of osteoclastogenesis by gp130-utilizing cytokines or interleukin-1 but not 1,25-dihydroxyvitamin D3 or parathyroid hormone. *J. Biol. Chem.* **274**, 19301–19308 (1999).
- Itoh, S. *et al.* A critical role for interleukin-6 family-mediated Stat3 activation in osteoblast differentiation and bone formation. *Bone* **39**, 505–512 (2006).
- Akira, S. *et al.* Molecular cloning of APRF, a novel IFN-stimulated gene factor 3 p91-related transcription factor involved in the gp130-mediated signaling pathway. *Cell* **77**, 63–71 (1994).
- Zhong, Z., Wen, Z. & Darnell, J. E. Jr. Stat3: a STAT family member activated by tyrosine phosphorylation in response to epidermal growth factor and interleukin-6. *Science* **264**, 95–98 (1994).
- Li, W., Liang, X., Kellendonk, C., Poli, V. & Taub, R. STAT3 contributes to the mitogenic response of hepatocytes during liver regeneration. *J. Biol. Chem.* **277**, 28411–28417 (2002).
- Robinson, D. S., Larche, M. & Durham, S. R. Tregs and allergic disease. *J. Clin. Invest.* **114**, 1389–1397 (2004).
- Yang, X. O. *et al.* STAT3 regulates cytokine-mediated generation of inflammatory helper T cells. *J. Biol. Chem.* **282**, 9358–9363 (2007).
- Happel, K. I. *et al.* Divergent roles of IL-23 and IL-12 in host defense against *Klebsiella pneumoniae*. *J. Exp. Med.* **202**, 761–769 (2005).
- Wolk, K. *et al.* IL-22 increases the innate immunity of tissues. *Immunity* **21**, 241–254 (2004).

28. Minegishi, Y. & Conley, M. E. Negative selection at the pre-BCR checkpoint elicited by human μ heavy chains with unusual CDR3 regions. *Immunity* **14**, 631–641 (2001).
29. Minegishi, Y. *et al.* Mutations in Ig α (CD79a) result in a complete block in B-cell development. *J. Clin. Invest.* **104**, 1115–1121 (1999).
30. Minegishi, Y. *et al.* An essential role for BLNK in human B cell development. *Science* **286**, 1954–1957 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We appreciate the willingness of the patients and the families to participate in this research study. This work is supported by the Japanese

Ministry of Education, Culture, Sports, Science and Technology, and the Japanese Ministry of Health, Labor and Welfare.

Author Contributions Y.M. designed and conducted most of the experiments; M.S. conducted the genetic analysis and the generation of osteoclasts; S.T., I.T., H.T., T.H., N.K., T.A., S.P. and A.M. diagnosed and collected samples; O.S. collected samples; H.K. oversaw the entire project; Y.M. and H.K. wrote the manuscript with comments from all co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Y.M. (yminegishi.mbch@tmd.ac.jp).

METHODS

Patients. An immunological work-up revealed high serum IgE in all the patients and eosinophilia in five of them. All other laboratory data examined were within the normal range, including the lymphocyte subpopulations, their proliferative responses to mitogens, the levels and subclasses of serum immunoglobulins, the oxidative burst of granulocytes, and the number and size of platelets.

Antibodies and cytokines. Antibodies against STAT3, tyrosine-phosphorylated STAT3, Flag and HA, and HRP-conjugated rabbit anti-mouse and goat anti-rabbit antibodies were purchased from Cell Signaling. The CD117 monoclonal antibody was from BD-PharMingen, and the CD3 monoclonal antibody (OKT3) was from Janssen Pharmaceutical. Recombinant human IL-6, IL-10, IL-12, IFN α , and GM-CSF were purchased from Peprotech, recombinant human IL-2 from Shionogi, and lipopolysaccharide (055:B5) from Sigma-Aldrich.

Isolation and culture of T cells and macrophages from PBMCs. Isolation and cell culture of T cells and macrophages were performed as described previously⁸. All the experiments were performed at least three times with three different controls.

Stimulation of cells with cytokines, and measurement of cytokines and IgM production. Cells were stimulated for the indicated time in culture with IL-6 (100 ng ml⁻¹), IL-10 (100 ng ml⁻¹), IL-12 (10 ng ml⁻¹), or IFN α (5 ng ml⁻¹) as described previously⁸.

RT-PCR and direct sequence analysis. Sequencing was performed with an ABI Prism dRhodamine Terminator kit and analysed with an ABI Prism 310 DNA Sequencer (Perkin-Elmer Applied Biosystems). At least two independent PCR products were sequenced.

Enzyme-linked DNA–protein interaction assay. Thirty micrograms of nuclear extracts were incubated in a 96-well microplate precoated with oligonucleotides encoding the consensus binding sequence for STAT1 or that for STAT3. Bound STAT3 or STAT1 was detected with specific antibodies plus an HRP-conjugated secondary antibody.

Retroviral infections. Retroviral infections were done with the retroviral vector pMX-IRES-GFP (a gift from T. Kitamura) carrying the wild-type or one of each mutant STAT3 sequence as described previously⁸.

STAT3 knock-down. Transfection of siRNA oligonucleotides was performed by using Lipofectamine-RNAiMAX reagent (Invitrogen). Forty hours after transfection, the cells were treated with IFN α (10 ng ml⁻¹) for 5 h.

Luciferase reporter assay. The reporter construct contained 4 repeated STAT3-responsive elements linked to a luciferase reporter gene. HeLa cells or HepG2 cells were transfected with the pcDNA3 vector bearing wild-type or mutant STAT3, the reporter construct, and an expression vector for *Renilla* luciferase driven by the CMV reporter, with FuGENE6 (Roche). The relative luciferase activity was determined by normalizing the values against the *Renilla* luciferase signal.

Tip60 is a haplo-insufficient tumour suppressor required for an oncogene-induced DNA damage response

Chiara Gorrini^{1*}, Massimo Squatrito^{1*†}, Chiara Luise², Nelofer Syed³, Daniele Perna¹, Landon Wark⁴, Francesca Martinato¹, Domenico Sardella¹, Alessandro Verrecchia¹, Samantha Bennett¹, Stefano Confalonieri², Matteo Cesaroni¹, Francesco Marchesi⁵, Milena Gasco⁶, Eugenio Scanziani⁵, Maria Capra², Sabine Mai⁴, Paolo Nuciforo², Tim Crook³, John Lough⁷ & Bruno Amati¹

The acetyl-transferase Tip60 might influence tumorigenesis in multiple ways¹. First, Tip60 is a co-regulator of transcription factors that either promote or suppress tumorigenesis, such as Myc and p53^{2–4}. Second, Tip60 modulates DNA-damage response (DDR) signalling¹, and a DDR triggered by oncogenes can counteract tumour progression^{5,6}. Using E_{μ} -myc transgenic mice that are heterozygous for a Tip60 gene (*Htatip*) knockout allele (hereafter denoted as *Tip60*^{+/-} mice), we show that Tip60 counteracts Myc-induced lymphomagenesis in a haplo-insufficient manner and in a time window that is restricted to a pre- or early-tumoral stage. Tip60 heterozygosity severely impaired the Myc-induced DDR^{7–9} but caused no general DDR defect in B cells. Myc- and p53-dependent transcription were not affected, and neither were Myc-induced proliferation, activation of the ARF–p53 tumour suppressor pathway or the resulting apoptotic response^{10–13}. We found that the human TIP60 gene (*HTATIP*) is a frequent target for mono-allelic loss in human lymphomas and head-and-neck and mammary carcinomas, with concomitant reduction in mRNA levels. Immunohistochemical analysis also demonstrated loss of nuclear TIP60 staining in mammary carcinomas. These events correlated with disease grade and frequently concurred with mutation of p53. Thus, in both mouse and human, Tip60 has a haplo-insufficient tumour suppressor activity that is independent from—but not contradictory with—its role within the ARF–p53 pathway^{1–3,14–16}. We suggest that this is because critical levels of Tip60 are required for mounting an oncogene-induced DDR in incipient tumour cells^{5,6}, the failure of which might synergize with p53 mutation towards tumour progression^{17–20}.

Heterozygous mice bearing a null allele of *Htatip* (hereafter denoted as *Tip60*^{+/-} mice) are viable and show no developmental or tumour-prone phenotype, whereas homozygous embryos that lack Tip60 die before implantation (M.-S. Kim and J.L., unpublished data). Breeding of *Tip60*^{+/-} mice with E_{μ} -myc transgenic mice revealed accelerated onset and enhanced penetrance of Myc-induced B-cell lymphomas in the *Tip60*^{+/-} background (Fig. 1a) with similar pathological features as in *Tip60*^{+/+} controls (Supplementary Fig. 1). Lymphomas that arose in heterozygous mice did not lose the wild-type *Tip60* allele but unexpectedly, instead, duplicated it at the

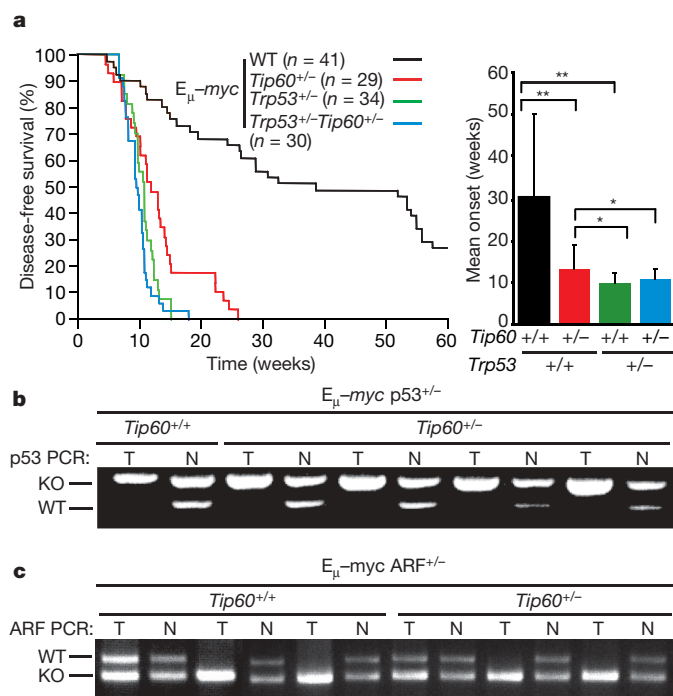


Figure 1 | Tip60 haplo-insufficiency accelerates Myc-induced lymphomagenesis but does not alleviate the pressure for p53 inactivation. **a**, Kaplan-Meier curves showing disease-free survival of E_{μ} -myc mice of the indicated genotypes. The bar graph reports disease onset in the same groups as average \pm s.d. (** $P \leq 0.001$, * $P \leq 0.05$). **b,c**, Genomic DNA from tumour (T) and normal (N) tail tissue from mice of the indicated genotypes was used for PCR with primers amplifying the wild-type (WT) and knockout (KO) alleles of the gene for p53 (*Trp53*; **b**) or ARF (**c**), as indicated. All tumours in E_{μ} -myc *Trp53*^{+/-} *Tip60*^{+/-} mice (10/10) underwent *Trp53* LOH (**b** and data not shown). As shown in **a**, these tumours developed slightly faster than in E_{μ} -myc *Tip60*^{+/-} mice, but at the same rate as in E_{μ} -myc *Trp53*^{+/-} mice. Similarly, *Arf* LOH (**c**)¹⁰ and tumour onset (data not shown) were equivalent in *Arf*^{+/-} and *Arf*^{+/-} *Tip60*^{+/-} backgrounds.

¹Department of Experimental Oncology, European Institute of Oncology (IEO), ²FIRC Institute of Molecular Oncology (IFOM), IFOM-IEO Campus, Milan 20139, Italy. ³Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, Chester Beatty Laboratories, London, SW3 6JB, UK. ⁴Manitoba Institute of Cell Biology and The Genomic Center for Cancer, Research and Diagnosis, CancerCare Manitoba, University of Manitoba, Winnipeg, Manitoba R3E 0V9, Canada. ⁵Department of Veterinary Pathology, Hygiene and Public Health, Section of Veterinary and Avian Pathology, University of Milan, Milan 20133, Italy. ⁶Department of Medical Oncology, San Croce e Carle Hospital, Cuneo 12100, Italy. ⁷Medical College of Wisconsin, Milwaukee, Wisconsin 53226, USA. [†]Present address: Department of Cancer Biology and Genetics, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA.

*These authors contributed equally to this work.

expense of the *neo*-targeted null allele, regaining wild-type expression of Tip60 (Supplementary Fig. 2a–d). Hence, Tip60 acts as a haplo-insufficient tumour suppressor at a pre- or early tumoral stage in E_{μ} -*myc* mice.

Heterozygosity for Tip60 did not significantly affect the homeostasis of pre-tumoral B cells, including Myc-induced proliferation, apoptosis (Fig. 2a–c) and the balance between pre-B ($B220^{+}IgM^{-}$) and mature B cells ($B220^{+}IgM^{+}$)²¹ (Fig. 2d). However, Tip60 heterozygosity severely impaired a Myc-induced DDR in B cells, as judged by immunostaining of phosphorylated ATM (ataxia telangiectasia mutated homologue) (Fig. 2e) and immunoblot analysis of phosphorylated forms of p53, checkpoint kinase 1 (Chk1), the histone variant H2AX (known as γ H2AX) and the Bcl-2 family-member Bid (Fig. 2f)²². Mouse embryo fibroblasts (MEFs) were isolated to confirm these observations *in vitro*, but MEFs expressed higher levels of Tip60 messenger RNA than did B cells (Supplementary Fig. 2e) and showed no haplo-insufficient phenotype (data not shown). Hence, we used lentiviral small-hairpin RNA (shRNA) vectors, which reduced Tip60 levels in MEFs and impaired the phosphorylation of Bid and H2AX after activation of a Myc-estrogen receptor chimera (MycER) (Supplementary Fig. 3a,b), showing that there is a cell-autonomous requirement of Tip60 for Myc-induced DDR.

In contrast to Myc, whole-body ionizing radiation (IR) induced an equivalent DDR in $Tip60^{+/+}$ and $Tip60^{+/-}$ B cells (Fig. 2g). Similarly, ultraviolet (UV)- or IR-induced DDR still occurred after Tip60 knockdown in MEFs (Supplementary Fig. 3c). Mutations in genes that regulate DDR accelerate spontaneous tumour onset in p53-null mice (for example, refs 17–20), but Tip60 haplo-insufficiency induced no significant acceleration in this context (Supplementary Fig. 4). In summary, $Tip60^{+/-}$ mice have a subtle defect in DDR signalling, which is only revealed—and associated with tumour acceleration—in a sensitized E_{μ} -*myc* background.

One mediator of Myc-induced DDR might have been ARF, the product of the alternative reading frame of the *Arf-ink4a* locus, which bound Tip60 and induced a Tip60-dependent DDR in transfected cells²³. However, the Myc-induced DDR was not reduced in *Arf*^{-/-} MEFs or pre-tumoral B cells (Supplementary Fig. 3d,e). Another likely mediator was ATM, whose activation might require Tip60 activity²⁴ and whose deletion enhances Myc-induced skin tumorigenesis⁸ and lymphomagenesis^{9,25}. However, we were unable to confirm that ATM and Tip60 interacted²⁴. Moreover, although the IR-induced DDR was impaired in *Atm*-null MEFs, the Myc-induced DDR was intact (Supplementary Fig. 3f). Therefore, if involved, ATM must be redundant for this response, perhaps with the related kinases ATR (ataxia-telangiectasia related) or DNA-PK (DNA-dependent protein kinase)^{22,26}. Tip60 modulates many other signalling pathways²⁷ some of which, like the p38 mitogen-activated protein kinase (MAPK) pathway, are involved in oncogene-induced stress (P. Sun, personal communication). Myc itself might trigger multiple genotoxic signals, including reactive oxygen species, and replicative or telomeric stress^{7,9,28}. Consistent with the latter, Myc slightly increased the frequency of aberrant (interstitial) telomeres in B cells independent of the *Tip60* genotype (data not shown). Thus, the Myc- and Tip60-dependent DDR is likely to be a complex multi-factorial event, the molecular basis of which remains to be unravelled.

An oncogene-induced DDR is thought to suppress tumorigenesis by activating p53 (refs 5, 6). Oncogenes also induce expression of ARF, which suppresses Mdm2-dependent p53 degradation²⁹. In E_{μ} -*myc* mice, this generates a strong selective pressure for loss of ARF or p53 function during tumour progression^{10,11}. Lymphomas arising in mice that are heterozygous for *p53* or *Arf* generally undergo loss-of-heterozygosity (LOH, that is, loss of the wild-type allele), and mutations that impair ARF–p53 function or block apoptosis alleviate the selective pressure for LOH^{11–13}. By contrast, LOH for *p53* or *Arf* still

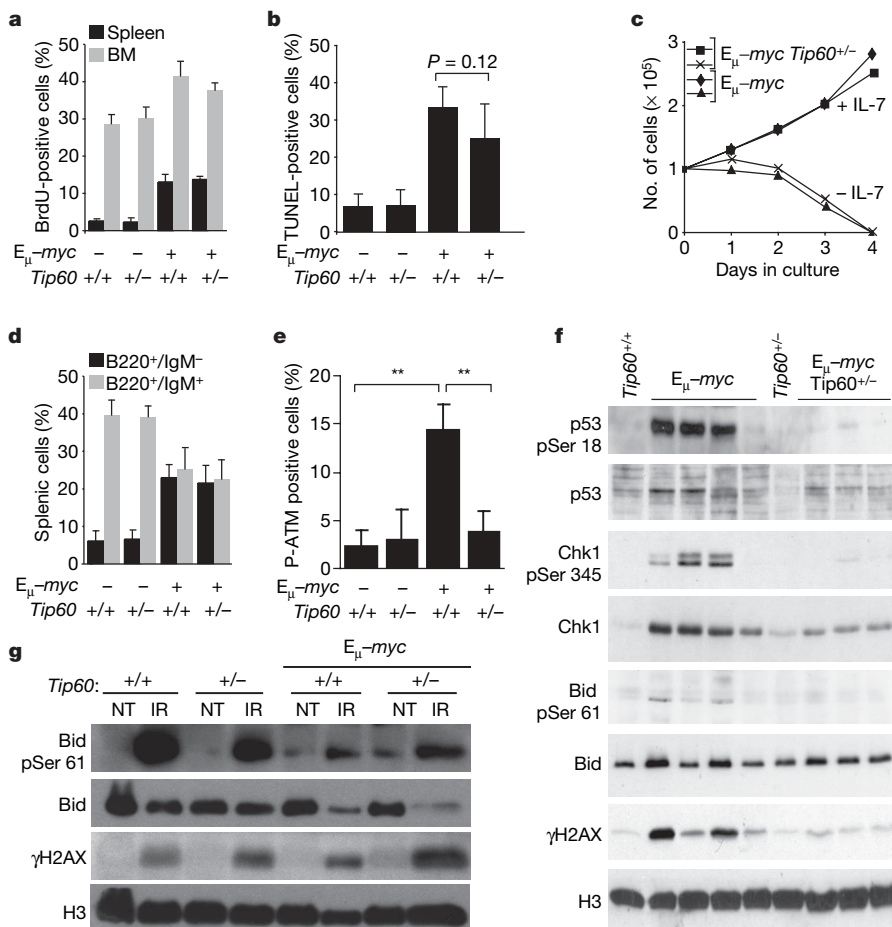


Figure 2 | Tip60 haplo-insufficiency does not affect B-cell homeostasis, but abrogates a Myc-induced DDR.

$B220^{+}$ cells were isolated from healthy young mice (4–6 weeks old) and analysed without any *ex vivo* culture, except for **c**. Bar graphs report average values \pm s.d. **a**, *In vivo* 5-bromodeoxyuridine (BrdU) incorporation in $B220^{+}$ -cells from spleen (black bars) and bone marrow (grey bars) of the indicated mice ($n = 3$ per genotype). **b**, TdT-mediated dUTP nick end labelling (TUNEL) staining of splenic $B220^{+}$ -cells ($n = 4$). A slight decrease in apoptosis was observed in the $Tip60^{+/-}$ background, but this remained below statistical significance **c**, *Ex vivo* growth curves of bone marrow-derived $B220^{+}$ cells in the presence or absence of interleukin (IL)-7. Transgenic B cells of either *Tip60* genotype showed similar rates of expansion and death *in vitro* in the presence and absence of IL-7, respectively¹⁰. **d**, Flow-cytometric analysis of splenic pre-B cells ($B220^{+}IgM^{-}$, black bars) or mature B cells ($B220^{+}IgM^{+}$, grey bars; $n = 5$). In E_{μ} -*myc* mice, expansion of pre-B cells is offset by increased apoptosis of mature B cells²¹. **e**, Immunofluorescence analysis of Ser 1981-phosphorylated ATM (P-ATM) in splenic $B220^{+}$ cells. The percentage of P-ATM⁺ cells is indicated as mean \pm s.d. (** $P < 0.01$, $n = 5$). This assay does not exclude recognition of additional phospho-epitopes besides P-ATM. **f**, Immunoblotting for total or phosphorylated forms of p53, Chk1, Bid, H2AX and histone H3 in splenic $B220^{+}$ cells. Each lane represents a different individual. **g**, Splenic $B220^{+}$ cells were isolated 2 h after whole-body IR (IR, 2 Gy) and immunoblotted as indicated. NT, not treated.

occurred in the *Tip60*^{+/-} background (Fig. 1b,c). Consistent with the notion that p53 remained functional in pre-tumoral *E_μ-myc Tip60*^{+/-} B cells, p53 levels were enhanced in these cells, even though phosphorylation of p53 Ser 18 was compromised (Fig. 2f). Finally, immunoblotting for members of the p53 and Bcl-2 families in lymphomas showed no significant difference between *Tip60*^{+/-} and *Tip60*^{+/-} mice (Supplementary Fig. 5), as would have been expected if those pathways had been differentially mutated. In summary, ARF is predominantly responsible for the induction of p53 activity and apoptosis in *E_μ-myc* mice. We infer that Tip60 (and the Myc-induced DDR) must suppress lymphomagenesis through additional mechanisms.

Tip60 is thought to contribute to DNA binding and transcriptional activation by transcription factors such as Myc or p53^{1-4,15}. In transgenic B cells, Tip60 haplo-insufficiency affected neither Myc-regulated gene expression (Fig. 3a, Supplementary Fig. 6a), nor binding of Myc to its target promoters for *Ncl* (nucleolin) and *Cad* (carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase) (Fig. 3b). Recruitment of Tip60 to *Ncl* was also unaffected⁴ (Fig. 3c). Analysis of p53-target genes revealed no Myc-dependent changes under steady-state conditions, but strong induction following whole-body IR (Fig. 3d, Supplementary Fig. 6b) accompanied by binding of p53 to target promoters (Fig. 3e), all independent of the *Tip60* genotype. The Bcl-2 family members Puma (also known as Bcl-2 binding component, encoded by *Bbc3*) and Bax (Bcl-2-associated X protein) are crucial mediators of apoptosis and tumour suppression in *E_μ-myc* mice^{12,30}, and their activation in human cells requires acetylation of p53 by Tip60 or the related histone acetyltransferase MOF (also known as *Myst1*)^{2,3}. As induction of Puma and Bax (Fig. 3d,e; Supplementary Fig. 6b) and apoptosis induced by irradiation (data not shown) still occurred in *Tip60*^{+/-} mice, this regulatory mechanism seems to be intact. In summary, Tip60 haplo-insufficiency affected neither Myc- nor p53-dependent transcription.

mRNA expression profiling showed that pre-tumoral B cells of *Tip60*^{+/-} and *Tip60*^{+/-} backgrounds clustered in distinct groups, based on differential hybridization of 1,041/19,906 (5%) probe sets. However, these differences were marginal and only a minority scored as significant by quantitative PCR (Supplementary Fig. 7). These subtle alterations in transcriptional profiles might have contributed to the DDR defect and faster tumour onset in *E_μ-myc Tip60*^{+/-} mice, although the critical target loci—if any—remain to be identified.

Given our data in the mouse, we surveyed the status of TIP60 in primary human tumours. *HTATIP* mRNA levels were reduced relative to matched normal tissue in 21/52 cases of ductal breast carcinoma, including 17/20 high-grade tumours (Supplementary Table 1; Fig. 4a). TIP60 expression was also reduced in head and neck squamous cell carcinoma (HNSCC) and low-grade B-cell lymphomas (follicular lymphoma, diffuse large B-cell lymphoma (DLBCL), Mantle cell lymphoma; Fig. 4b) but not high-grade Burkitt's lymphomas (data not shown). The latter finding might indicate that TIP60 has a rate-limiting function in these MYC-induced tumours, perhaps consistent with the *Tip60* duplication that is seen in *E_μ-myc Tip60*^{+/-} mice (Supplementary Fig. 2a, b). In all tumour types, low levels of *HTATIP* mRNA almost systematically correlated with LOH at single nucleotide polymorphisms (SNPs) within the *HTATIP* locus (Fig. 4a,b). In breast carcinomas and HNSCC we detected no aberrant CpG methylation of the *HTATIP* promoter. Only in few cases of DLBCL and follicular lymphoma did we find concomitant SNP LOH and CpG methylation (Supplementary Table 1). In summary, loss of one *HTATIP* allele is a frequent event in the human tumours analysed, whereas epigenetic silencing of the second allele is rare. Thus, as in mice, TIP60 behaves as a haplo-insufficient tumour suppressor in humans.

We surveyed TIP60 protein levels in 10 tumour types by immunohistochemistry (IHC) on tissue microarrays. Nuclear TIP60 staining was reduced in a fraction of breast, colon, lung and stomach

carcinomas (Supplementary Fig. 8). Most significantly, although virtually all normal breast epithelia showed nuclear TIP60, this was lost in 39% of tumours. Analysis of a larger cohort reinforced this trend, with loss of nuclear TIP60 in 72% (129/179) of breast carcinomas (Fig. 4d), and even more when considering high-grade tumours

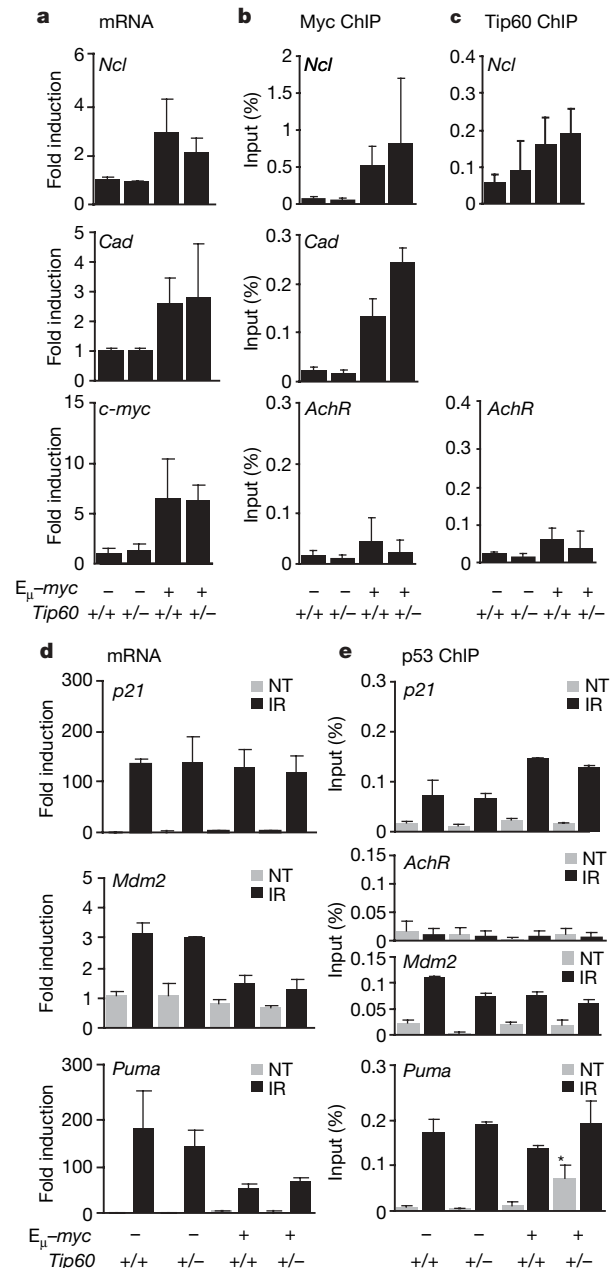


Figure 3 | Tip60 haplo-insufficiency does not affect Myc- and p53-dependent transcription. Splenic B220⁺ cells from 4–6-week-old mice were used in all experiments. **a**, Quantitative RT-PCR analysis of mRNA levels for the Myc-target genes *Ncl* and *Cad*, and for *c-Myc*, normalized to ribosomal protein 9 (*rps9*) mRNA. The genotypes of the mice are indicated at the bottom ($n = 3$ and 5 for non-transgenic and transgenic groups, respectively). **b**, Chromatin immunoprecipitation (ChIP) analysis of Myc binding to *Ncl* (intron 1) and *Cad* (promoter). The muscarinic acetylcholine receptor M1 gene (*AchR*, also known as *Chrm1*) was used as a non-binding control⁴ ($n \geq 5$). **c**, ChIP analysis of Tip60 binding to *Ncl* and *AchR* ($n \geq 5$). **d**, mRNA analysis as in **a** for cyclin-dependent kinase inhibitor 1a (*Cdkn1a*, or *p21*), *mdm2* and *Puma* mRNAs. Mice were either treated with whole-body irradiation (IR: 2 Gy) 2 h before isolation of the cells, or left untreated (NT) ($n = 3$). **e**, ChIP analysis of p53 binding to the *Cdkn1a*, *Mdm2* and *Puma* promoters in young mice treated as in **d**. The asterisk indicates a high background value on *Puma* in this sample, which was present also in the no-antibody control ChIP (not shown). All graphs report mean \pm s.d.

(Fig. 4e) or those that were stained by the K_i -67 antibody (Fig. 4f). Most carcinomas *in situ* (CIS) also showed reduced levels of nuclear TIP60, suggesting that downregulation of TIP60 had already been selected in these early lesions.

Besides the strong nuclear staining, we observed a fainter, but specific, cytoplasmic signal (Fig. 4h, Supplementary Fig. 9). Nuclear and cytoplasmic TIP60 were not correlated, although an inverse (but non-significant) trend was observed when scoring cytoplasmic TIP60 against the aforementioned pathological features (Supplementary Table 2). This indicates that, besides downregulation, delocalization of TIP60 (or of some TIP60 isoform) might also contribute to disease progression. Clarification of this issue awaits functional and mechanistic understanding of TIP60 trafficking.

Finally, we addressed the status of p53 in relationship to TIP60. In both breast carcinoma and HNSCC, the frequency of TIP60 LOH was higher in the subset of samples with p53 mutations (Fig. 4c). Similarly, loss of nuclear TIP60 staining by IHC was more frequent in breast tumours with strong (that is, mutant) p53 staining (Fig. 4g,h).

Thus, downregulation of TIP60 was most frequently associated with mutant rather than wild-type p53, as would have been expected if TIP60 acted mainly through the p53 pathway.

We have shown that TIP60 suppresses tumorigenesis in a haplo-insufficient and p53-independent manner in humans and mice. This does not contradict the idea that TIP60 is required for p53 function^{1-3,14-16}, but reveals a different tumour suppressor activity, which is most sensitive to sub-threshold levels of TIP60. The mouse data suggest that this activity is related to the role of TIP60 in the oncogene-induced DDR, supporting the concept that DDR signaling in pre-tumoral lesions serves as a barrier to tumour progression^{5,6,8}, but also implying that this barrier does not always rest on p53. This might result from the predominance of the ARF-MDM2 pathway for p53 activation, as happens in E_{μ} -myc mice^{10,11}. The positive association between TIP60 and p53 alterations in human tumours further indicates that these proteins act on parallel pathways, loss of which might synergize during tumour progression. A possible mechanism is suggested by studies on mice with combined

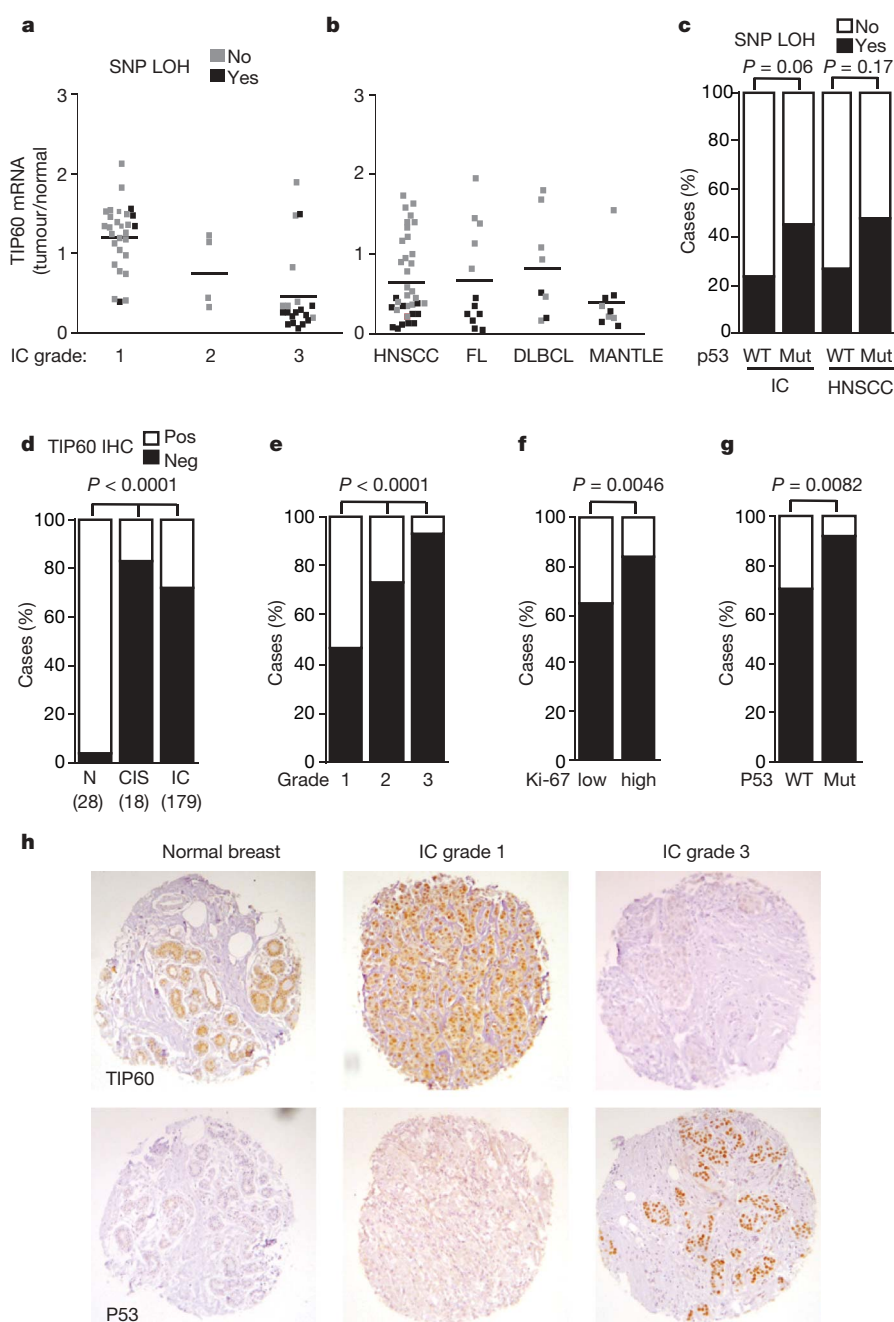


Figure 4 | TIP60 is a haplo-insufficient tumour suppressor in human tumours.

a, Scatter plot depicting *HTATIP* mRNA expression in breast tumours of different grades, relative to matched normal tissue. Samples with and without SNP LOH at the *HTATIP* locus are coded as indicated. The horizontal lines indicate the average expression in each group. **b**, Same as **a** in head and neck squamous carcinoma (HNSCC), follicular lymphoma (FL), diffuse large B-cell lymphoma (DLBCL) and mantle cell lymphoma (MANTLE). **c**, Cases from breast carcinoma or HNSCC with or without LOH were grouped by P53 status based on genomic sequencing. WT, wild type; Mut, mutant. **d**, Positive (POS) or negative (NEG) TIP60 nuclear staining by IHC in normal breast tissue (N), carcinoma *in situ* (CIS) or invasive breast carcinomas (IC). **e–g**, Invasive breast carcinomas were plotted as in **d** but grouped by tumour grade (**e**), K_i -67 staining (**f**), or P53 staining (**g**). P53 was scored as mutant when the staining was positive. **h**, Representative images of TIP60 and P53 staining by IHC in the indicated breast tissues. The data for **a–c** and **d–g** are given in Supplementary Tables 1 and 2, respectively. In all graphs, *P* values from Fisher's test are shown.

DDR and p53 deficiencies, in which chromosomal translocations cause accelerated tumour onset (for example, refs 18–20). Failure of the Myc-induced DDR in E_{μ} -myc *Tip60*^{+/-} mice, combined with loss of p53 function (which occurs independently owing to selective pressure against apoptosis), might similarly enhance genomic instability, adding to that caused by Myc itself and favouring the selection of genomic alterations in incipient tumour clones. Besides Tip60, other gene products and pathways that modulate the oncogene-induced DDR are likely to contribute to tumour suppression: the E_{μ} -myc model offers an ideal platform for their genetic dissection.

METHODS SUMMARY

Mice were monitored twice a week for lymphoma development by peripheral lymph node palpation. Tumours and lymphoid organs were dissected and either prepared for histology or processed for molecular analysis. Pre-tumoral B cells were purified from the spleen or bone marrow of young (4–6 weeks) healthy mice and used for flow-cytometry, immunofluorescence, chromatin immunoprecipitation, mRNA analysis or protein analysis. MEFs were expanded under low oxygen tension and infected at early passages with the indicated MycER- and/or shRNA-expressing viruses. MycER was activated by addition of 4-hydroxy-tamoxifen (OHT) to the culture medium for the indicated time. For molecular genetic and immuno-histochemical analysis of human tumours, tissues were obtained with informed patient consent and with local ethical committee approval. For detailed experimental procedures, see Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 19 April; accepted 22 June 2007.

- Squatrito, M., Gorrini, C. & Amati, B. Tip60 in DNA damage response and growth control: many tricks in one HAT. *Trends Cell Biol.* **16**, 433–442 (2006).
- Tang, Y., Luo, J., Zhang, W. & Gu, W. Tip60-dependent acetylation of p53 modulates the decision between cell-cycle arrest and apoptosis. *Mol. Cell* **24**, 827–839 (2006).
- Sykes, S. M. *et al.* Acetylation of the p53 DNA-binding domain regulates apoptosis induction. *Mol. Cell* **24**, 841–851 (2006).
- Frank, S. R. *et al.* Myc recruits the Tip60 histone acetyl-transferase complex to chromatin. *EMBO Rep.* **4**, 575–580 (2003).
- Bartkova, J. *et al.* DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature* **434**, 864–870 (2005).
- Gorgoulis, V. G. *et al.* Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature* **434**, 907–913 (2005).
- Prochownik, E. V. & Li, Y. The ever expanding role for c-Myc in promoting genomic instability. *Cell Cycle* **6**, 1024–1029 (2007).
- Pusapati, R. V. *et al.* ATM promotes apoptosis and suppresses tumorigenesis in response to Myc. *Proc. Natl Acad. Sci. USA* **103**, 1446–1451 (2006).
- Reimann, M. *et al.* The Myc-evoked DNA damage response accounts for treatment resistance in primary lymphomas *in vivo*. *Blood* 11 June 2007 (doi: 10.1182/blood-2007-02-075614).
- Eischen, C. M., Weber, J. D., Roussel, M. F., Sherr, C. J. & Cleveland, J. L. Disruption of the ARF-Mdm2-p53 tumor suppressor pathway in Myc-induced lymphomagenesis. *Genes Dev.* **13**, 2658–2669 (1999).
- Schmitt, C. A., McCurrach, M. E., de Stanchina, E., Wallace-Brodeur, R. R. & Lowe, S. W. INK4a/ARF mutations accelerate lymphomagenesis and promote chemoresistance by disabling p53. *Genes Dev.* **13**, 2670–2677 (1999).
- Eischen, C. M., Roussel, M. F., Korsmeyer, S. J. & Cleveland, J. L. Bax loss impairs Myc-induced apoptosis and circumvents the selection of p53 mutations during Myc-mediated lymphomagenesis. *Mol. Cell. Biol.* **21**, 7653–7662 (2001).
- Schmitt, C. A. *et al.* Dissecting p53 tumor suppressor functions *in vivo*. *Cancer Cell* **1**, 289–298 (2002).
- Berns, K. *et al.* A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**, 431–437 (2004).
- Tyteca, S., Vandromme, M., Legube, G., Chevillard-Briet, M. & Trouche, D. Tip60 and p400 are both required for UV-induced apoptosis but play antagonistic roles in cell cycle progression. *EMBO J.* **25**, 1680–1689 (2006).
- Legube, G. *et al.* Role of the histone acetyl transferase Tip60 in the p53 pathway. *J. Biol. Chem.* **279**, 44825–44833 (2004).
- Westphal, C. H. *et al.* atm and p53 cooperate in apoptosis and suppression of tumorigenesis, but not in resistance to acute radiation toxicity. *Nature Genet.* **16**, 397–401 (1997).
- Celeste, A. *et al.* H2AX haploinsufficiency modifies genomic stability and tumor susceptibility. *Cell* **114**, 371–383 (2003).
- Bassing, C. H. *et al.* Histone H2AX: a dosage-dependent suppressor of oncogenic translocations and tumors. *Cell* **114**, 359–370 (2003).
- Morales, J. C. *et al.* 53BP1 and p53 synergize to suppress genomic instability and lymphomagenesis. *Proc. Natl Acad. Sci. USA* **103**, 3310–3315 (2006).
- Nilsson, J. A. *et al.* Targeting ornithine decarboxylase in Myc-induced lymphomagenesis prevents tumor formation. *Cancer Cell* **7**, 433–444 (2005).
- Sancar, A., Lindsey-Boltz, L. A., Unsal-Kacmaz, K. & Linn, S. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu. Rev. Biochem.* **73**, 39–85 (2004).
- Eymen, B. *et al.* p14ARF activates a Tip60-dependent and p53-independent ATM/ATR/CHK pathway in response to genotoxic stress. *Mol. Cell. Biol.* **26**, 4339–4350 (2006).
- Sun, Y., Jiang, X., Chen, S., Fernandes, N. & Price, B. D. A role for the Tip60 histone acetyltransferase in the acetylation and activation of ATM. *Proc. Natl Acad. Sci. USA* **102**, 13182–13187 (2005).
- Shreeram, S. *et al.* Regulation of ATM/p53-dependent suppression of myc-induced lymphomas by Wip1 phosphatase. *J. Exp. Med.* **203**, 2793–2799 (2006).
- Jiang, X., Sun, Y., Chen, S., Roy, K. & Price, B. D. The FATC domains of PIKK proteins are functionally equivalent and participate in the Tip60-dependent activation of DNA-PKcs and ATM. *J. Biol. Chem.* **281**, 15741–15746 (2006).
- Lehner, B., Crombie, C., Tischler, J., Fortunato, A. & Fraser, A. G. Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. *Nature Genet.* **38**, 896–903 (2006).
- Dominguez-Sola, D. *et al.* Non-transcriptional control of DNA replication by c-Myc. *Nature* **448**, 445–451 (2007).
- Sharpless, N. E. INK4a/ARF: a multifunctional tumor suppressor locus. *Mutat. Res.* **576**, 22–38 (2005).
- Hemann, M. T. *et al.* Suppression of tumorigenesis by the p53 target PUMA. *Proc. Natl Acad. Sci. USA* **101**, 9333–9338 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Campaner, G. Natoli, G. Del Sal, M. Foiani, F. d'Adda di Fagnana, E. Belloni and H. Müller for discussions and comments, F. Contegno for laboratory set-up and management, A. Gobbi, M. Capillo and B. Giulini for management of mouse colonies, L. Tizzoni and L. Bernard for qPCR, I. Muradore for FACS analysis, J. Cleveland and J.-C. Marine for E_{μ} -myc and p53 knockout mice, A. Gross for phospho-Bid antibodies and P.-G. Pelicci for his continuous support. This work was supported by the Italian Association for Cancer Research (AIRC) and the 'Ricerca Finalizzata' program of the Italian Health Ministry (to B.A.), in part by the NIH (to J.L.) and by the National Cancer Institute of Canada and CancerCare Manitoba Foundation (to S.M.).

Author Contributions C.G., M.S. and B.A. conceived the work and designed the experiments. B.A. supervised the project and wrote the manuscript. C.G. and M.S. performed all of the experimental work on mice and cells, with the exceptions listed below. F. Martinato optimized the ChIP assays on freshly isolated B-cells. S.B. characterized *Tip60*^{+/-} MEFs. D.P. performed the experiments in ARF and ATM knockout MEFs. D.S. performed the genotyping of all mice. S.C. and M. Cesaroni contributed bioinformatic analysis. L.W. and S.M. performed cytogenetic analysis of telomeres in B cells. F. Marchesi and E.S. performed the mouse pathology. A.V. constructed and optimized the *Tip60* shRNA and provided technical assistance. N.S., M.G. and T.C. contributed the genetic analysis of human tumours. C.L., M. Capra and P.N. produced and analysed the IHC data on human tumours. J.L. constructed the *Tip60*^{+/-} mice.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.A. (bruno.amati@ifom-ileo-campus.it).

METHODS

Mice and genotyping. E_{μ} -myc transgenic mice³¹, and $Tip60^{+/-}$, $Trp53^{+/-}$ or $Arf^{+/-}$ mice^{32,33} and their offspring were bred to obtain the genotypic combinations described in this paper (all in 129Sv-C57BL/6 mixed background). $Tip60^{+/-}$ mice bear one null allele, lacking *Htatip* exons 1–9. These mice are born at mendelian ratios, are viable into adulthood without any manifest phenotype and are fertile. Homozygous $Tip60^{-/-}$ embryos die at pre-implantation stages (M-S. Kim and J.L., unpublished results). Mice were monitored twice a week for signs of morbidity and tumour development. In E_{μ} -myc mice, tumours were scored by peripheral lymph node palpation¹³. Sick animals were killed, and tumours and lymphoid organs were analysed by histology or processed for further analysis. The $Tip60$ knockout and PCR-based genotyping strategies will be described elsewhere (M-S. Kim and J.L. submitted). The PCR primers for amplification of *Htatip* from genomic DNA were: *tip1* GCTAGCCAGTAA-GCATCATT and AGCTTGTGTGCCATCATCTC; *tip2* TGAGCGTGAAGGACATCAGTG and GTCACCCATCCAGACGT; *tip3* AAAGTGAGACAGGCGAAGA and CTCCTAGGCCTCTTCTCTCTC. *Trp53* and *Arf* genotyping was as described^{32,33}. The PCR primers for genotyping of E_{μ} -myc transgenic mice were GGTTTAATGAATTGAAGTTGCCA and TTCTTGCCCTGCGTATATCAGTC, based on the E_{μ} -myc DNA junction (accession no. M12344)^{31,34}. Where appropriate, 3–4-week-old mice were exposed to 2 Gy ionizing radiation from a ¹³⁷Cs source.

Flow cytometry, magnetic sorting of B cells, cell culture and gene transfer. We obtained single-cell suspensions by pressing the spleen through nylon cell strainers or by leg bone crushing and subsequent hypotonic lysis of red blood cells. For pre-tumoral analysis of B-cell differentiation status, splenic cells were incubated for 1 h with anti-CD45R/B220 biotin-conjugated antibodies (BD Biosciences Pharmingen), fluorescein isothiocyanate (FITC)-conjugated anti-IgM (Jackson ImmunoResearch), and, after washing, resuspended in phycoerythrin (PE)-conjugated streptavidin (BD Biosciences Pharmingen). We measured the proliferation rates of B cells from spleen and bone marrow using a Flow Kit (BD Biosciences Pharmingen) as described by the manufacturer. Briefly, animals were injected intraperitoneally with BrdU (40 µg per g body weight) in sterile PBS, and bone marrow and spleen were harvested 2 h later. We incubated 10⁶ cells with antibodies against B220 (PE-conjugated) and BrdU (FITC-conjugated). All samples were analysed by FACSCalibur (Becton Dickinson). To isolate B cells, we incubated single-cell suspensions from bone marrow or spleen with B220 MicroBeads (Miltenyi Biotec) and enriched them by magnetic cell sorting (MACS), according to the manufacturer's instructions (Miltenyi Biotec). Primary bone marrow-derived pre-B cells were cultured in RPMI 1640 medium supplemented with 10% fetal calf serum (hyClone), 2 mM glutamine, 55 µM 2-mercaptoethanol, with or without recombinant murine IL-7 (50 ng ml⁻¹) (PeproTech Inc.). Splenic B cells were maintained in the presence of 20 ng ml⁻¹ mouse recombinant IL-4 (PeproTech Inc.) and 20 µg ml⁻¹ lipopolysaccharide (Sigma). MEFs were obtained from embryonic day (E)12.5 embryos, grown for two-population doublings (P1) and then viably frozen. MEFs were cultured in Dulbecco-modified Eagle medium (DMEM) supplemented with 10% fetal calf serum (hyClone), 2 mM glutamine, and 55 µM 2-mercaptoethanol under low-oxygen tension (3%). MEFs were infected with pBabe-Puro or pBabe-Puro-mycER retroviruses and selected in puromycin-containing medium. MycER was induced by stimulation with 4-hydroxy-tamoxifen (OHT) for 24 or 48 h. For $Tip60$ knock-down experiments, MEFs were infected with a Lentilox-Puro lentiviral vector (modified from Lentilox-GFP)³⁵ expressing shRNAs complementary to the *Tip60* mRNA. The target sequences were GGACGAAGCGGAAATCTA (*Tip 1*) and GGGCCAGTATATCCTAACT (*Tip 2*). Infectious retro- or lentiviruses were produced by transfection into Phoenix or 293T packaging cell lines. Where appropriate, MEFs were exposed to UV (20 J m⁻²) or γ -radiation (15 Gy) and harvested 2 h later.

Immunoblot analysis, immunofluorescence and TUNEL assay. For direct immunoblot analysis, freshly MACS-sorted splenic B cells, lymphomas or cultured MEFs were lysed in Laemmli sample buffer, containing 62.5 mM Tris-HCl pH 6.8, 2% (w/v) SDS, 5% (v/v) 2-mercaptoethanol and 10% (v/v) glycerol. This stringent lysis condition is essential for quantitative and unbiased recovery of $Tip60$ (see legend to Supplementary Fig. 2). For the analysis shown in Supplementary Fig. 5, lymphomas were solubilized in 50 mM Hepes pH 7.4, 150 mM NaCl, 1.5 mM EDTA, 10% (v/v) glycerol, 1% (v/v) NP40. Proteins (20–40 µg per lane) were separated by SDS-PAGE and analysed by immunoblotting with primary antibodies against $Tip60^4$, p53 (from P.G. Pelicci), p53 pSer 18 (9284, Cell Signalling), p19ARF (AB80, Abcam), vinculin (V9131, Sigma), Chk1 pSer 345 (2341, Cell Signalling), γ H2AX (05-636, Upstate), histone H3 (9715, Cell Signalling), Bid pSer 61 (from A. Gross³⁶ and A300-527A, Bethyl), Bid (AF860, R&D Systems), Bim (AAP-330, Stressgen), Bcl-XL (AAM-080, Stressgen), actin (AC-40, Sigma) and Bcl2 (M0887, Dako Cytomation). Other

antibodies were from Santa Cruz: Mdm2 (C-18), Puma (N-19), Mcl1 (S-19), Chk1 (G-4) and tubulin (H-235). For phospho-ATM immunofluorescence, MACS-sorted B cells were layered on poly-lysine-coated glass slides, fixed with 4% w/v paraformaldehyde, permeabilized in 0.1% Triton X-100 in PBS, and incubated for 1 h at room temperature with anti-ATM phospho-Ser 1981 antibodies (200–301-400, Rockland), then with Cy3-conjugated goat-anti-rabbit antibodies (Jackson ImmunoResearch). Nuclei were counterstained with DAPI. We counted 500–1,000 cells per individual. For apoptosis measurements by TUNEL assay, we stained B cells on glass slides with the *In situ* Cell Death Detection kit (Roche) according to the manufacturer's instructions.

mRNA analysis and chromatin immunoprecipitation. Total RNA was prepared using Trizol reagent (Invitrogen) and further purified onto RNeasy columns (Qiagen). Complementary DNA (cDNA) was produced using the reverse-transcriptase SuperscriptII (Invitrogen). Quantitative chromatin immunoprecipitation (ChIP) was performed as described³⁷, with only the following adjustment for *in vivo* analysis: MACS-sorted splenic B cells ($\geq 20 \times 10^6$) were resuspended in PBS with 0.5% BSA and 2 mM EDTA at room temperature and fixed by addition of 1% formaldehyde for 10 min. Fixation was stopped by addition of 0.125 M glycine. Cells were washed three times in PBS and 2 mM EDTA, resuspended in SDS buffer and processed for ChIP³⁷. Real-time PCR reactions on ChIP-derived genomic DNA or cDNA were performed with either SYBR Green Reaction Mix, or TaqMan Gene expression assays (Applied Biosystems), the latter with primer/probe combinations based on the Universal Probe Library Set (Roche). PCR primer sequences for the SYBR Green reactions were the following: *Ncl* forward: GCGTGCTGACTCCACGT; *Ncl* reverse: CGAAATCACCTCTTAAAGCAGCA; *Cad* forward: GGCTGCTTGCGCCGT; *Cad* reverse: CAGAGCCCTAAGCGTAGTGAG; *Puma* forward: CCGTTAGTCTGAGCGTACTCC; *Puma* reverse: CGCTTGACACACTGACACT; *Mdm2* forward: GGTCAAGTTGGGACAGCTG; *Mdm2* reverse: AGCGTTTAAATAACCCAGCTG; *Cip1* (*p21*) forward: CCTTTCTATCAGCCCCAGAGGATA; *Cip1* (*p21*) reverse: GGGACATCCTTAATTA-TCTGGGGT; *AchR* forward: AGTGCCCCCTGCTGTCACT; *AchR* reverse: CCCTTCTCTGGTCCCAAGA.

Microarray hybridization and data analysis. Total RNA was prepared using Trizol reagent (Invitrogen) and further purified onto RNeasy columns (Qiagen). RNA quality was analysed with the Agilent 2100 Bioanalyser (Agilent Technologies). We amplified 500 ng of total RNA using the T7-polymerase-based double linear amplification protocol³⁸. We hybridized 25 µg of cRNA probe onto the GeneChip Mouse Genome 430 2.0 Array ChipSet (Affymetrix), and processed it according to Affymetrix technical protocols. The average intensity of every array was scaled to a predefined value (target intensity) of 500, to make arrays comparable. The expression profile data pre-processed with MAS5 were exported to GeneSpring GX 7.3 (Agilent Technologies) for further elaboration. According to the GeneSpring normalization procedure, the 50th percentile of all measurements was used as a positive control, within each hybridization array, and each measurement for each gene was divided by this control. The bottom 10th percentile was used for background subtraction. Among different hybridization arrays, each gene was divided by the median of its measurements in all samples. Data were then log transformed for subsequent analysis. We prefiltered the expression data by taking into account both MAS5 'Absolute Call' flags and average expression measurements within each group. Briefly, from a total of 45,121 probe sets, we selected only those called present or marginal (P or M) at least once across all samples, and whose mean raw expression levels were ≥ 200 within at least one class of samples (E_{μ} -myc $Tip60^{+/-}$ or E_{μ} -myc $Tip60^{+/-}$). The prefiltering method removed those probe sets whose expression signal was constantly too close to the background signal throughout the entire set of samples (probably defined as 'Absent' by the MAS5 algorithm and consequently not expressed). This procedure allowed us to select 19,906 probe sets. To find genes whose expression levels significantly differed between pre-tumoral B cells from E_{μ} -myc $Tip60^{+/-}$ or E_{μ} -myc $Tip60^{+/-}$ mice, we adopted a supervised method of analysis, using the GeneSpring software. We calculated mean values among individual experimental groups for each probe set, and derived fold-change ratios between E_{μ} -myc $Tip60^{+/-}$ and E_{μ} -myc $Tip60^{+/-}$ samples. A difference of twofold cut-off was applied to select upregulated and downregulated genes. The statistical analysis was performed with GeneSpring using Welch's approximate *t*-test, with a *P* value cut-off of 0.05, without the assumption of equality of variances. We used the Benjamini and Hochberg false discovery rate (FDR) for multiple testing correction³⁹. By this analysis, 1,041 probe sets were found to be significantly regulated between the two classes of samples.

Molecular genetic and immunohistochemical analysis of primary human tumours. For molecular genetic analysis, squamous and breast carcinomas were as described^{40,41}. We obtained normal and neoplastic tissue from multiple 5- or 10-µm tissue sections for each case by micro-dissection or from peripheral blood

lymphocytes, and isolated nucleic acids using Qiagen reagents as described⁴¹. The sequence of the p53 gene *TRP53* was analysed by amplification and individual sequencing of multiple plasmid clones. We analysed the expression of *HTATIP* by quantitative RT-PCR using SYBR Green Reaction Mix and *GAPDH* as a control gene. Cancers were classified as having downregulation of *HTATIP* mRNA if the cancer:normal signal ratio was <0.5 . Primer sequences for *HTATIP* were: forward: TGACCAAGTGTGACCTACGA; reverse primer: CTGGGATATACTCTTGTCTTACG; product size: 104 bp.

To analyse LOH, we identified SNPs with a reasonable percentage of heterozygosity in European populations from the NCBI SNP browser. We analysed the following SNPs in the *HTATIP* gene: rs573589, rs4645927, rs2236682 and rs535111.

We looked for aberrant CpG methylation using methylation-specific PCR (MSP). Genomic DNA was modified as described⁴². Three primer pairs were used, each of which maps to distinct regions of a CpG island containing 59 CpG dinucleotides at the 5' end of the *HTATIP* gene, at positions 65,235,926–65,236,600 of chromosome 11 (<http://genome.ucsc.edu/cgi-bin/hgGateway>). Primer sequences were as follows: pair 1, methylated, forward: TTTTTCGATAAAGTTAAAAACG; reverse: CATAACGTACACAACCAATAACGAT; product size: 126 bp. Unmethylated, forward: TTTTTCGATAAAGTTAAAAATGG; reverse: CCCATAACATACACAACCAATAACA; product size: 127 bp. Pair 2, methylated, forward: TTTTTCGATAAAGTTAAAAACGG; reverse: CATAACGTACACAACCAATAACGAT; product size: 125 bp. Unmethylated, forward: TTTTTCGATAAAGTTAAAAATGG; reverse: CCCATAACATACACAACCAATAACA; product size: 127 bp. Pair 3, methylated, forward: GTTGTTTTTTAGGGGGAGATAATC; reverse: CTCGAAACAATAAAAAACACG; product size: 230 bp. Unmethylated, forward: GTTGTTTTTTAGGGGGAGATAATT; reverse: CTCTCAAAACAATAAAAAACACAC; product size: 232 bp.

We carried out bisulphite sequencing using three primer pairs to independently map the majority of the *HTATIP* CpG island. Bisulphite pair 1, forward: GTTGTTTTTTAGGGGGAGATAAT; reverse: CTAAACCTCCTTCAATCACAAATAC; product size: 284 bp; CpGs in product: 19. Bisulphite pair 2, forward: GTGTTTTTTAAAGGTTTTTTT; reverse: AACTCTACCCACCTCCCCTAAC; product size: 182 bp; CpGs in product: 12. Bisulphite pair 3, forward: AATGAAGTTATAGATTGGGGGAG; reverse: AACCTAAATAAAAAACCCATAAC; product size: 266 bp; CpGs in product: 22.

For the large-scale immuno-histochemical screening, we used formalin-fixed and paraffin-embedded specimens derived from various normal and tumour tissues, indicated in Supplementary Fig. 8. Tip60 analysis in breast cancer progression was performed on an independent data set constituted of 18 carcinomas *in situ* (CIS) and 179 primary invasive carcinomas (IC). In invasive carcinoma, we also analysed correlations between Tip60, Ki-67, and p53 staining and clinicopathological features. Tissue microarrays (TMA) were specifically designed for the screening and prepared as described⁴³. Briefly, two representative normal (when available) and tumour areas (diameter, 0.6 mm) from each sample,

identified previously on haematoxylin and eosin stained sections, were removed from the donor blocks and deposited on the recipient block using a custom-built precision instrument (Tissue Arrayer, Beecher Instruments). Sections (2 μ m) of the resulting recipient block were cut, mounted on glass slides, and processed for immunohistochemistry using the affinity-purified anti-Tip60 RLPV antibody⁴ diluted 1:200 and incubated overnight. Tumour samples showing Tip60 nuclear staining in any cell nucleus were considered positive. Sections were also incubated for 2 h at room temperature using anti-p53 antibody (DO1, IFOM Monoclonal facility) diluted 1:5000 and Ki-67 antibody (MIB1, Dako) diluted 1:200. Cut-offs of $\geq 10\%$ and $\geq 20\%$ of positive cells were used to establish p53 protein accumulation and high Ki-67 expression, respectively.

Statistical analysis. All the statistical analysis was done by two-tailed Student's *t*-test, except for data from molecular genetic and immunohistochemistry on human tumours, where Fisher's exact test was used. Correlations of Tip60 with clinical, histological, and biological parameters in breast cancer were assessed using the Pearson Chi-square.

31. Adams, J. M. *et al.* The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice. *Nature* **318**, 533–538 (1985).
32. Jacks, T. *et al.* Tumor spectrum analysis in p53-mutant mice. *Curr. Biol.* **4**, 1–7 (1994).
33. Kamijo, T. *et al.* Loss of the ARF tumor suppressor reverses premature replicative arrest but not radiation hypersensitivity arising from disabled atm function. *Cancer Res.* **59**, 2464–2469 (1999).
34. Corcoran, L. M., Cory, S. & Adams, J. M. Transposition of the immunoglobulin heavy chain enhancer to the myc oncogene in a murine plasmacytoma. *Cell* **40**, 71–79 (1985).
35. Rubinson, D. A. *et al.* A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference. *Nature Genet.* **33**, 401–406 (2003).
36. Kamei, I. *et al.* Proapoptotic BID is an ATM effector in the DNA-damage response. *Cell* **122**, 593–603 (2005).
37. Frank, S. R., Schroeder, M., Fernandez, P., Taubert, S. & Amati, B. Binding of c-Myc to chromatin mediates mitogen-induced acetylation of histone H4 and gene activation. *Genes Dev.* **15**, 2069–2082 (2001).
38. Eberwine, J. *et al.* Analysis of gene expression in single live neurons. *Proc. Natl Acad. Sci. USA* **89**, 3010–3014 (1992).
39. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B* **57**, 289–300 (1995).
40. Bergamaschi, D. *et al.* p53 polymorphism influences response in cancer chemotherapy via modulation of p73-dependent apoptosis. *Cancer Cell* **3**, 387–402 (2003).
41. Crichton, D. *et al.* DRAM, a p53-induced modulator of autophagy, is critical for apoptosis. *Cell* **126**, 121–134 (2006).
42. Syed, N. *et al.* Transcriptional silencing of Polo-like kinase 2 (SNK/PLK2) is a frequent event in B-cell malignancies. *Blood* **107**, 250–256 (2006).
43. Kononen, J. *et al.* Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nature Med.* **4**, 844–847 (1998).

LETTERS

Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1

Eros Lazzerini Denchi¹ & Titia de Lange¹

When telomeres are rendered dysfunctional through replicative attrition of the telomeric DNA or by inhibition of shelterin¹, cells show the hallmarks of ataxia telangiectasia mutated (ATM) kinase signalling^{2–4}. In addition, dysfunctional telomeres might induce an ATM-independent pathway, such as ataxia telangiectasia and Rad3-related (ATR) kinase signalling, as indicated by the phosphorylation of the ATR target CHK1 in senescent cells^{2,5} and the response of ATM-deficient cells to telomere dysfunction^{6,7}. However, because telomere attrition is accompanied by secondary DNA damage, it has remained unclear whether there is an ATM-independent pathway for the detection of damaged telomeres. Here we show that damaged mammalian telomeres can activate both ATM and ATR and address the mechanism by which the shelterin complex represses these two important DNA damage signalling pathways. We analysed the telomere damage response on depletion of either or both of the shelterin proteins telomeric repeat binding factor 2 (TRF2) and protection of telomeres 1 (POT1) from cells lacking ATM and/or ATR kinase signalling. The data indicate that TRF2 and POT1 act independently to repress these two DNA damage response pathways. TRF2 represses ATM, whereas POT1 prevents activation of ATR. Unexpectedly, we found that either ATM or ATR signalling is required for efficient non-homologous end-joining of dysfunctional telomeres. The results reveal how mammalian telomeres use multiple mechanisms to avoid DNA damage surveillance and provide an explanation for the induction of replicative senescence and genome instability by shortened telomeres.

The shelterin complex can be compromised through inhibition of TRF2, which, together with TRF1, anchors this complex onto the double-stranded telomeric DNA¹. Shelterin also contains POT1, a single-stranded DNA-binding protein that is tethered to TRF1 and TRF2 through the POT1 binding partner TPP1 (previously TINT1, PTP, PIP1). When TRF2 is compromised, telomeres elicit a robust DNA damage signal, which can be detected by their decoration with DNA damage factors, forming 'telomere dysfunction-induced foci'⁴ (TIFs). The ATM kinase is activated^{3,4,8}, leading to cell cycle arrest mediated by the tumour protein p53 and the cyclin-dependent kinase inhibitor p21 (ref. 9). To determine whether TRF2 loss can induce a DNA damage response in the absence of ATM, we generated *Atm*^{−/−} mouse embryonic fibroblasts (MEFs)¹⁰ carrying the conditional allele of the gene for TRF2 (*Terf2*^{fl/fl}). Cells were immortalized with SV40 large T antigen to bypass the cell cycle arrest that normally occurs in response to telomere damage. As shown previously, deletion of TRF2 from *Atm*^{+/+} cells leads to the induction of TIFs (Fig. 1a, b and Supplementary Fig. 1a) and phosphorylation of Chk2 (Fig. 1c). In contrast, in *Atm*^{−/−} cells, the formation of TIFs (Fig. 1a, b and Supplementary Fig. 1a) and phosphorylation of Chk2 (Fig. 1c) in response to TRF2 deletion were largely abrogated. The deletion of TRF2 from *Atm*^{−/−} cells was confirmed by PCR (Supplementary Fig.

1b), immunofluorescence (Supplementary Fig. 1c), disappearance of repressor activator protein 1 (Rap1; Fig. 1c), a TRF2-interacting protein whose stability depends on TRF2 (refs 8, 11), and telomeric chromatin immunoprecipitation (ChIP; Supplementary Fig. 2).

The inability of *Atm*^{−/−} cells to respond to TRF2 deletion indicated that the ATR kinase is not involved in the detection of this type of telomere dysfunction. Consistent with the absence of an ATR response, TRF2 deletion did not induce phosphorylation of Chk1 (Fig. 1d). However, recent data have indicated that ATR activation at double-stranded DNA breaks (DSBs) induced by irradiation might depend on ATM¹², raising the possibility that the ATR response to dysfunctional telomeres might also require ATM. To test the role of ATR further, we used a short hairpin RNA (shRNA) that diminished ATR protein levels and compromised ATR signalling as deduced from the lack of ultraviolet-induced Chk1 phosphorylation (Supplementary Fig. 3). This level of ATR inhibition had no effect on TIF formation or the phosphorylation of Chk2 after TRF2 deletion from *Atm*^{+/+} cells (Supplementary Fig. 3a, b). We therefore conclude that the telomere damage generated by TRF2 loss primarily activates ATM, not ATR.

Unexpectedly, *Atm*^{−/−} cells showed a 15-fold reduction in the rate of telomere fusion in response to deletion of TRF2 (Fig. 1e and Supplementary Fig. 4a). The effect of ATM status on non-homologous end joining (NHEJ) of dysfunctional telomeres was also detectable in genomic blots (Supplementary Fig. 4b). Whereas *Terf2*^{−/−}*Atm*^{+/+} cells showed the expected appearance of large-molecular-weight fragments that are typical of fused telomeres, *Terf2*^{−/−}*Atm*^{−/−} cells revealed a largely unchanged telomeric restriction fragment pattern. Furthermore, the 3' telomeric overhang signal was unaffected by deletion of TRF2 from *Atm*^{−/−} cells (Supplementary Fig. 4b). In contrast, *Terf2*^{−/−}*Atm*^{+/+} cells showed a significant reduction in 3' overhang signals, as expected from the tight link between 3' overhang processing and NHEJ^{8,13}. The role of ATM in the NHEJ of dysfunctional telomeres was verified by *in vivo* deletion of TRF2 from quiescent mouse hepatocytes using an *Mx*-Cre system¹¹. In this setting, ATM deficiency also diminished the occurrence of the telomere fusions (Supplementary Fig. 5). We conclude that, in primary cells as well as in immortalized cells, ATM is required for the recognition and processing of dysfunctional telomeres generated by loss of TRF2.

These results raised the possibility that the ATR kinase is repressed by another component of shelterin. Because the activation of the ATR kinase involves the binding of the replication protein RPA to single-stranded DNA¹⁴, we considered the possibility that POT1 could be this factor. In this regard, *Saccharomyces cerevisiae* cell division cycle-13 (Cdc13), a protein with similarities to POT1, represses signalling by mitosis entry checkpoint-1 (Mec1)^{15–17}. To test the role of POT1 in repression of the ATR kinase, we used POT1-null MEFs. Rodents have two POT1 genes, *Pot1a* and *Pot1b*, both of which are required for telomere function¹⁸. Deletion of *Pot1a*, alone or with *Pot1b*,

¹Laboratory for Cell Biology and Genetics, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA.

induces TIFs^{18,19}. In contrast to the results with TRF2, the TIF response in *Pot1a/b* double knockout cells was significantly reduced when we inhibited ATR with shRNA (Fig. 2a, b). Furthermore, unlike TRF2-null cells, *Pot1a/b* double knockout cells showed phosphorylation of Chk1 as well as Chk2, consistent with ATR signalling, and the

phosphorylation of these downstream kinases depended on ATR (Fig. 2c). The role of ATR in the DNA damage response elicited by loss of POT1a was corroborated by conditional deletion of ATR from MEFs²⁰ (Fig. 2d). Efficient ATR deletion was confirmed on the basis of diminished induction of Chk1 phosphorylation by ultraviolet irradiation (Fig. 2e). We used two approaches to remove POT1a from the telomeres of these cells: an shRNA that was previously shown to affect POT1a specifically¹⁸, and an shRNA that effectively knocks down TPP1 as monitored by ChIP²¹. Both approaches resulted in the formation of TIFs and, in both cases, Cre-mediated deletion of ATR significantly suppressed the telomere damage response (Fig. 2d). In contrast, the status of ATM expression in the cells did not affect the response to the type of telomere damage induced by POT1a or TPP1 knockdown (Supplementary Fig. 6a–c). To test whether POT1 represses ATR signalling at telomeres of human cells, we used a colon carcinoma cell line (HCT116) with a conditional *ATR* allele²². The cells were infected with a *POT1* shRNA that has been shown to elicit a telomere damage response²³.

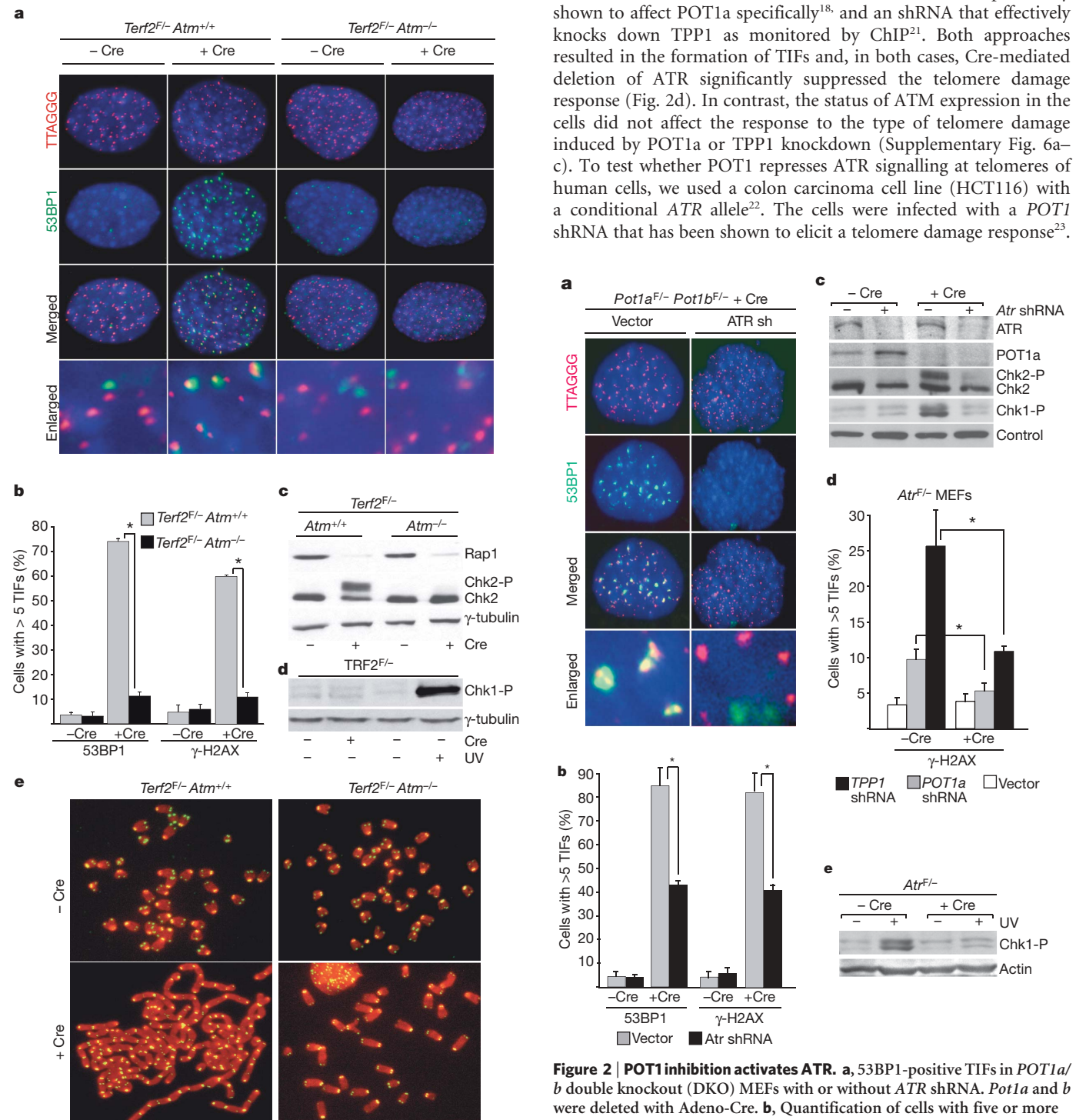


Figure 1 | TRF2 deletion activates ATM, not ATR. **a**, 53BP1-positive TIFs in MEFs of the indicated genotypes (DNA stained with DAPI (blue)). **b**, Quantification of 53BP1- and γH2AX-positive TIFs (mean and s.d. of triplicate experiments; $n \geq 150$; asterisk, $P < 0.001$ calculated using a two-tailed Student's t -test). **c**, **d**, Immunoblots showing Rap1 and Chk2 (**c**), and Chk1 phosphorylation induced by ultraviolet irradiation (25 J m^{-2}) but not TRF2 deletion (**d**). **e**, Metaphase spreads of MEFs of the indicated genotype and treatment were stained for telomeric DNA (green) and DAPI (red).

Figure 2 | POT1 inhibition activates ATR. **a**, 53BP1-positive TIFs in *POT1a/b* double knockout (DKO) MEFs with or without ATR shRNA. *Pot1a* and *b* were deleted with Adeno-Cre. **b**, Quantification of cells with five or more 53BP1-positive or γH2AX-positive TIFs (mean and s.d. of triplicate experiments; $n \geq 150$; asterisk, $P < 0.005$ based on a two-tailed Student's t -test). **c**, Immunoblotting for ATR, POT1a, Chk2 and Chk1-P in MEFs after the indicated treatment. **d**, TIFs in *Atr^{F/-}* MEFs expressing *Pot1a* sh3 or *Tpp1* sh3. TIFs were detected by FISH-immunofluorescence either before (-Cre) or 4 d after (+Cre) infection with pWZL-Cre. Cells with five or more γH2AX TIFs were scored as positive (error bars, s.e.m.; $n \geq 150$; asterisk, $P < 0.01$ based on a two-tailed Student's t -test). **e**, Immunoblotting for Chk1-P in *Atr^{F/-}* MEFs treated as indicated.

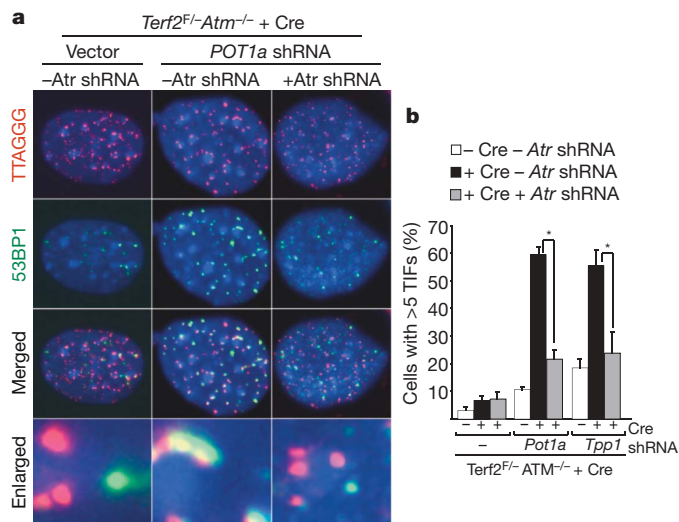


Figure 3 | ATR activation in *Terf2^{-/-} Atm^{-/-}* cells on inhibition of POT1a. *Atm^{-/-} Terf2^{F/-}* MEFs were infected with shRNAs to inhibit POT1a, TPP1 or ATR as indicated. TRF2 was deleted with a Cre adenovirus and TIFs were detected and quantified 5 d later as described in Fig. 1. **a**, **b**, Visualization and quantification of the effect of *Atr* shRNA on TIFs induced by inhibition of POT1a or TPP1 in *Atm^{-/-} Terf2^{F/-}* cells. Error bars, s.d.; $n \geq 150$; asterisk, $P < 0.01$ based on a two-tailed Student's *t*-test.

Consistent with the data obtained in mouse cells, the telomere damage response resulting from inhibition of human POT1 was largely dependent on ATR (Supplementary Fig. 7). Collectively, these data show that ATR signalling at mammalian telomeres is repressed by POT1 proteins.

The data described above predict that the lack of ATR activation in *Terf2^{-/-} Atm^{-/-}* MEFs is due to residual POT1a at the telomeres (Supplementary Fig. 2). To test this, we used shRNA to inhibit POT1a or TPP1 in *Terf2^{-/-} Atm^{-/-}* cells. As deletion of POT1b does not induce a DNA damage response¹⁸, we used a *Pot1b* shRNA as a negative control. We found that inhibition of POT1a or TPP1, but not POT1b, elicited a telomere damage response in *Terf2^{-/-} Atm^{-/-}* cells (Supplementary Fig. 8a, b). We next tested the involvement of ATR in this response using an *Atr* shRNA in conjunction with either *Pot1a* or *Tpp1* shRNA. Two selectable markers were used to ensure the presence of both shRNAs. The results indicated that ATR signalling contributes to TIF formation when POT1a or TPP1 are compromised in *Terf2^{-/-} Atm^{-/-}* cells (Fig. 3a, b). These data further confirm the role of POT1 in the repression of ATR and show that

the TRF2-independent recruitment of POT1 or TPP1 is sufficient to repress ATR activation.

Unexpectedly, inhibition of POT1a or TPP1 in *Terf2^{-/-} Atm^{-/-}* cells resulted in frequent telomere fusions (Fig. 4a, Supplementary Fig. 9). Telomere fusions occurred on 18 and 36% of chromosomes when cells were treated with *Pot1a* and *Tpp1* shRNAs, respectively. This is a significant increase from the background level of telomere fusions in *Terf2^{-/-} Atm^{-/-}* cells (4% of chromosomes; Fig. 4a, Supplementary Fig. 9a). The occurrence of telomere fusions was not anticipated as *Pot1a/b* double knockout cells have a very weak telomere fusion phenotype¹⁸. Knockdown of POT1a or TPP1 did not induce telomere fusions in *Atm^{-/-}* cells with normal TRF2 levels or in wild-type cells (Supplementary Fig. 9a), confirming previous findings that telomere fusions are effectively blocked by TRF2. The stimulation of telomere fusion by loss of POT1a from *Terf2^{-/-} Atm^{-/-}* cells was also detected in genomic blots (Supplementary Fig. 9b). Inhibition of TPP1 led to a more severe telomere fusion phenotype than did *Pot1a* shRNA. This could be due to more effective removal of POT1a from telomeres through TPP1, the additional removal of POT1b, or other effects of TPP1 inhibition. Because NHEJ of telomeres after TRF2 deletion involved signalling by ATM, we investigated whether ATR signalling contributes to the NHEJ pathway that joins telomeres on POT1a loss from *Terf2^{-/-} Atm^{-/-}* cells. Inhibition of ATR with shRNA partially suppressed the telomere fusions induced by inhibition of POT1a or TPP1 (Fig. 4a, Supplementary Fig. 9a).

Collectively, these data show that TRF2 and POT1 function independently to repress the activation of the ATM and ATR kinases at natural chromosome ends (Fig. 4b). How TRF2 prevents activation of the ATM kinase is not yet clear, although it has been suggested that TRF2, through its ability to interact with ATM, might block its activation²⁴. We speculate that the ATR pathway is inhibited by POT1 by blocking the binding of RPA to the single-stranded telomeric DNA. Although RPA is more abundant, POT1 could be an effective competitor because of its shelterin-mediated enrichment at telomeres. The repression of ATM and ATR by TRF2 and POT1 explains how cells detect critically shortened telomeres. Because the abundance of shelterin at telomeres depends on the length of the duplex telomeric repeat array, short telomeres contain less TRF2 and POT1 (ref. 25). This diminished loading of TRF2 and POT1 is expected ultimately to lead to derepression of the ATM and ATR kinases, resulting in cell cycle arrest and inappropriate DNA repair at telomeres.

The dissection of the DNA damage response at mammalian telomeres also sheds light on the interplay between the ATM and ATR kinases. We show that ATM is not required for the activation of ATR, because removal of POT1 leads to an ATR response in *Atm^{-/-}*

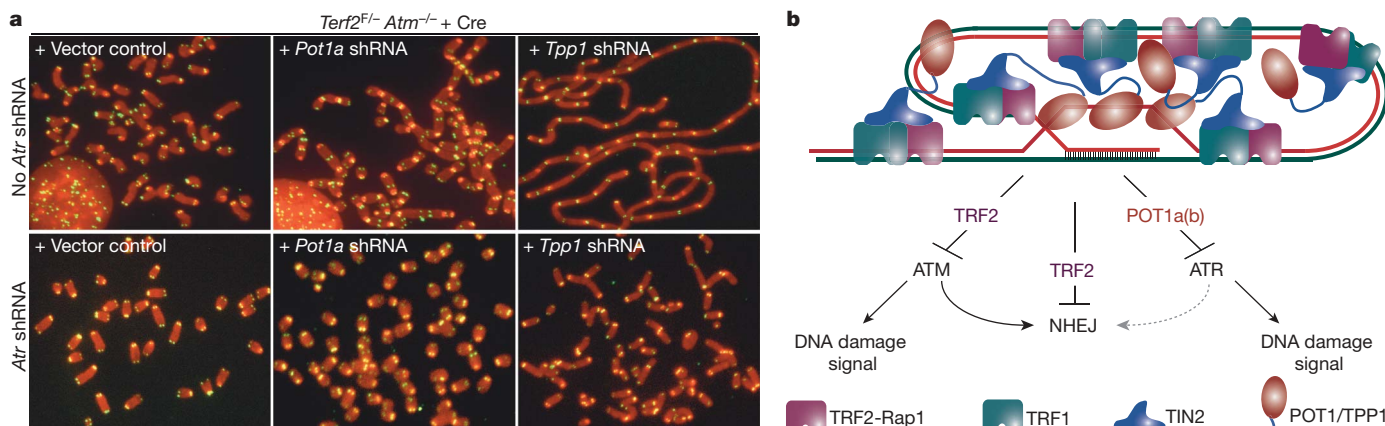


Figure 4 | Role of POT1 and ATR in NHEJ of telomeres lacking TRF2. *Terf2^{-/-} Atm^{-/-}* cells were treated with shRNAs to modulate the levels of POT1a, TPP1 and ATR as described in Fig. 3. **a**, Telomere fusions visualized in metaphase spreads stained for telomeric DNA (green) and DAPI (red).

cells. This finding contrasts with the situation at irradiation-induced DSBs, where ATM signalling can promote activation of the ATR kinase, possibly by stimulating the formation of single-stranded DNA at the broken end¹². At telomeres, which normally contain single-stranded DNA, such processing might not be necessary. Conversely, ATR is not required for the activation of ATM at telomeres. When TRF2 is removed from telomeres, ATM is activated without a contribution of ATR, which remains repressed by the residual POT1 at the chromosome ends.

Finally, our data reveal an unanticipated dependence of telomere fusions on signalling by either ATM or ATR (Fig. 4b). NHEJ of dysfunctional telomeres is primarily repressed by TRF2, possibly because the t-loop configuration blocks the end-loading of the Ku70/80 complex^{26,27}. However, when TRF2 is deleted in *Atm*^{-/-} cells, telomere NHEJ is inefficient. Whereas POT1 loss *per se* does not induce telomere fusions as long as TRF2 is functional, POT1 inhibition in *Trf2*^{-/-}*Atm*^{-/-} cells does. Our data indicate that the key event in this setting is the activation of ATR. We therefore suggest that NHEJ at telomeres requires both the loss of TRF2 and the activation of either ATM or ATR (Fig. 4b). It is possible that the dependencies described here for NHEJ at dysfunctional telomeres might also hold for other forms of DNA damage. The modest effect of ATM deficiency on NHEJ in other settings²⁸ might be due to compensation by ATR signalling. Because the ATM and ATR kinase pathways are controlled separately by TRF2 and POT1, telomeres provide a unique opportunity to manipulate these pathways and to uncover regulatory interactions that otherwise might remain obscured.

METHODS

MEFs from E13.5 embryos obtained from crosses between *Atm*^{+/-} (Jackson Labs) and *Trf2*^{fl-8} mice were immortalized at passage 2 with pBabeSV40LT (a gift from G. Hannon). Cre was introduced by retroviral infection using Hit&Run *Cre-GFP* or *pWZ-Cre* expression vectors or using adenoviral infections with the Ad5 CMV Cre virus (Resource Centre, The University of Iowa). *Mx*-Cre mediated deletion of TRF2 from quiescent hepatocytes was performed as described¹¹. Protein and genomic DNA were isolated 4 days after pI-pC induction of Cre.

ChIP^{25,29}, immunoblotting⁸, immunofluorescence⁸ and fluorescent *in situ* hybridization (FISH)/immunofluorescence co-staining⁵ were performed as described using the following antibodies: TRF2 (1254); Rap1 (1252); TRF1 (644); POT1a (1220); TPP1 (1550); Chk2 (BD biosciences); Phospho-Chk1 (Ser 345) (Cell Signaling); ATR (N-19) (Santa Cruz); 53BP1 antibody (Novus, NB 100-304); γ H2AX (Upstate Biotechnology); γ tubulin (Sigma); and actin (Santa Cruz). Cells with at least five telomeric 53BP1 (or γ H2AX) foci were scored as TIF positive; $n \geq 150$ for each experiment. Data reported are averages of three independent experiments.

shRNAs were expressed using a pSuperior-puromycin retroviral vector (OligoEngine) or (for ATR) the same vector modified to contain the hygromycin marker. Target sequences: *Atr* (sh3-1), GGAGATGCAACTCGTTTAA; *Tpp1* (sh3; ref. 21), GGACACATGGGCTGACGGA; *Pot1a* (sh3; ref. 18): GCATTTC-TCTACAACATTA.

Mouse telomeric DNA was analysed on CHEF gels⁸ and by FISH¹⁸ using previously described protocols.

Received 2 May; accepted 5 July 2007.

Published online 8 August 2007.

- de Lange, T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.* **19**, 2100–2110 (2005).
- d'Adda di Fagagna, F. *et al.* A DNA damage checkpoint response in telomere-initiated senescence. *Nature* **426**, 194–198 (2003).
- Karlseder, J., Broccoli, D., Dai, Y., Hardy, S. & de Lange, T. p53- and ATM-dependent apoptosis induced by telomeres lacking TRF2. *Science* **283**, 1321–1325 (1999).
- Takai, H., Smogorzewska, A. & de Lange, T. DNA damage foci at dysfunctional telomeres. *Curr. Biol.* **13**, 1549–1556 (2003).

- Herbig, U., Jobling, W. A., Chen, B. P., Chen, D. J. & Sedivy, J. M. Telomere shortening triggers senescence of human cells through a pathway involving ATM, p53, and p21(CIP1), but not p16(INK4a). *Mol. Cell* **14**, 501–513 (2004).
- Qi, L. *et al.* Short telomeres and ataxia-telangiectasia mutated deficiency cooperatively increase telomere dysfunction and suppress tumorigenesis. *Cancer Res.* **63**, 8188–8196 (2003).
- Wong, K. K. *et al.* Telomere dysfunction and *Atm* deficiency compromises organ homeostasis and accelerates ageing. *Nature* **421**, 643–648 (2003).
- Celli, G. B. & de Lange, T. DNA processing not required for ATM-mediated telomere damage response after TRF2 deletion. *Nature Cell Biol.* **7**, 712–718 (2005).
- Smogorzewska, A. & de Lange, T. Different telomere damage signaling pathways in human and mouse cells. *EMBO J.* **21**, 4338–4348 (2002).
- Barlow, C. *et al.* *Atm*-deficient mice: a paradigm of ataxia telangiectasia. *Cell* **86**, 159–171 (1996).
- Lazzerini Denchi, E., Celli, G. & de Lange, T. Hepatocytes with extensive telomere deprotection and fusion remain viable and regenerate liver mass through endoreduplication. *Genes Dev.* **20**, 2648–2653 (2006).
- Jazayeri, A. *et al.* ATM- and cell cycle-dependent regulation of ATR in response to DNA double-strand breaks. *Nature Cell Biol.* **8**, 37–45 (2006).
- Dimitrova, N. & de Lange, T. MDC1 accelerates nonhomologous end-joining of dysfunctional telomeres. *Genes Dev.* **20**, 3238–3243 (2006).
- Zou, L. & Elledge, S. J. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science* **300**, 1542–1548 (2003).
- Bertuch, A. A. & Lundblad, V. The maintenance and masking of chromosome termini. *Curr. Opin. Cell Biol.* **18**, 247–253 (2006).
- Enomoto, S., Glowczewski, L., Berman, J., MEC3, MEC1, and DDC2 are essential components of a telomere checkpoint pathway required for cell cycle arrest during senescence in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **13**, 2626–2638 (2002).
- Garvik, B., Carson, M. & Hartwell, L. Single-stranded DNA arising at telomeres in *cdc13* mutants may constitute a specific signal for the RAD9 checkpoint. *Mol. Cell Biol.* **15**, 6128–6138 (1995).
- Hockemeyer, D., Daniels, J. P., Takai, H. & de Lange, T. Recent expansion of the telomeric complex in rodents: Two distinct POT1 proteins protect mouse telomeres. *Cell* **126**, 63–77 (2006).
- Wu, L. *et al.* Pot1 deficiency initiates DNA damage checkpoint activation and aberrant homologous recombination at telomeres. *Cell* **126**, 49–62 (2006).
- Brown, E. J. & Baltimore, D. Essential and dispensable roles of ATR in cell cycle arrest and genome maintenance. *Genes Dev.* **17**, 615–628 (2003).
- Hockemeyer, D. *et al.* Telomere protection by mammalian POT1 requires interaction with TPP1. *Nature Struct. Mol. Biol.* advanced online publication, doi:10.1038/nsmb1270 (15 July 2007).
- Cortez, D., Guntuku, S., Qin, J. & Elledge, S. J. ATR and ATRIP: partners in checkpoint signaling. *Science* **294**, 1713–1716 (2001).
- Hockemeyer, D., Sfeir, A. J., Shay, J. W., Wright, W. E. & de Lange, T. POT1 protects telomeres from a transient DNA damage response and determines how human chromosomes end. *EMBO J.* **24**, 2667–2678 (2005).
- Karlseder, J. *et al.* The telomeric protein TRF2 binds the ATM kinase and can inhibit the ATM-dependent DNA damage response. *PLoS Biol.* **2**, E240 (2004).
- Loayza, D. & de Lange, T. POT1 as a terminal transducer of TRF1 telomere length control. *Nature* **423**, 1013–1018 (2003).
- Celli, G. B., Lazzerini Denchi, E. & de Lange, T. Ku70 stimulates fusion of dysfunctional telomeres yet protects chromosome ends from homologous recombination. *Nature Cell Biol.* **8**, 885–890 (2006).
- Griffith, J. D. *et al.* Mammalian telomeres end in a large duplex loop. *Cell* **97**, 503–514 (1999).
- Lobrich, M. & Jeggo, P. A. Harmonising the response to DSBs: a new string in the ATM bow. *DNA Repair (Amst.)* **4**, 749–759 (2005).
- Ye, J. Z. *et al.* POT1-interacting protein PIP1: a telomere length regulator that recruits POT1 to the TIN2/TRF1 complex. *Genes Dev.* **18**, 1649–1654 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. White for invaluable help in maintaining our mouse colonies, D. Argibay for assistance with genotyping, E. Brown for providing the ATR conditional knockout mice, H. Takai for technical advice and suggestions, D. Hockemeyer for providing *Pot1* double knockout cells and the *Tpp1* and *Pot1a* shRNA constructs, S. Soll for his contributions to the initial stages of this research and members of the de Lange laboratory for critical comments on the manuscript. This work was supported by a grant from the NIH.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.deL. (delange@mail.rockefeller.edu).

LETTERS

Crystal structure of the MgtE Mg^{2+} transporterMotoyuki Hattori¹, Yoshiki Tanaka¹, Shuya Fukai², Ryuichiro Ishitani¹ & Osamu Nureki^{1,3}

The magnesium ion Mg^{2+} is a vital element involved in numerous physiological processes. Mg^{2+} has the largest hydrated radius among all cations, whereas its ionic radius is the smallest. It remains obscure how Mg^{2+} transporters selectively recognize and dehydrate the large, fully hydrated Mg^{2+} cation for transport¹. Recently the crystal structures of the CorA Mg^{2+} transporter^{2–5} were reported^{6–8}. The MgtE family of Mg^{2+} transporters is ubiquitously distributed in all phylogenetic domains^{9–11}, and human homologues have been functionally characterized and suggested to be involved in magnesium homeostasis^{12–14}. However, the MgtE transporters have not been thoroughly characterized. Here we determine the crystal structures of the full-length *Thermus thermophilus* MgtE at 3.5 Å resolution, and of the cytosolic domain in the presence and absence of Mg^{2+} at 2.3 Å and 3.9 Å resolutions, respectively. The transporter adopts a homodimeric architecture, consisting of the carboxy-terminal five transmembrane domains and the amino-terminal cytosolic domains, which are composed of the superhelical N domain and tandemly repeated cystathionine-β-synthase domains. A solvent-accessible pore nearly traverses the transmembrane domains, with one potential Mg^{2+} bound to the conserved Asp 432 within the pore. The transmembrane (TM)5 helices from both subunits close the pore through interactions with the ‘connecting helices’, which connect the cystathionine-β-synthase and transmembrane domains. Four putative Mg^{2+} ions are bound at the interface between the connecting helices and the other domains, and this may lock the closed conformation of the pore. A structural comparison of the two states of the cytosolic domains showed the Mg^{2+} -dependent movement of the connecting helices, which might reorganize the transmembrane helices to open the pore. These findings suggest a homeostasis mechanism, in which Mg^{2+} bound between cytosolic domains regulates Mg^{2+} flux by sensing the intracellular Mg^{2+} concentration. Whether this presumed regulation controls gating of an ion channel or opening of a secondary active transporter remains to be determined.

The crystallographic asymmetric unit contains two oppositely oriented MgtE dimers (Supplementary Fig. 1), which are mostly identical, with a root mean square (r.m.s.) deviation of 0.513 Å for 884 Cα atoms. Therefore, we describe the structure of one dimer (molecules A and B) in the following discussion.

The homodimeric transporter has overall dimensions of $90 \times 65 \times 55$ Å³, consisting of the N-terminal cytosolic domains and the C-terminal transmembrane domain with five membrane-spanning helices per monomer (Fig. 1 and Supplementary Fig. 2). The two subunits symmetrically embrace each other, with both the cytosolic and transmembrane domains interacting tightly. The cytosolic domain consists of two subdomains: the N-terminal subdomain (N domain) forms a right-handed superhelix (ten helices per two turns; Fig. 1 and Supplementary Fig. 3), which shares structural similarity with the C-terminal domain of the human FancF protein¹⁵; the C-terminal subdomain is composed of typical duplicated CBS domains¹⁶ (Fig. 1 and Supplementary Fig. 3), which are known to

play a regulatory function in other transporters, such as human chloride channels and the osmoregulated ABC transporter OpuA^{16–18}. The cytosolic domain is acidic: 52 of the 244 residues are Asp and Glu, which are clustered in the N and CBS domain interface (Supplementary Fig. 4).

The linker region between the cytosolic and transmembrane domains contains a stretching helix—referred to as a ‘connecting helix’—which is oriented perpendicularly to the membrane interface and interacts with the TM5 helix as well as the TM2 and H4b helices (Figs 1 and 2a, b). These interactions would be based on the dipole moments of the connecting helix and TM5, and on the shape complementarity (van der Waals interaction) at the interface of the four helices (Fig. 1 and Supplementary Figs 5 and 6). The highly conserved acidic residues Asp 246, Asp 247, Asp 250, Glu 253, Glu 255, Glu 258 and Asp 259 are clustered on the connecting helix (Supplementary Fig. 7).

The transmembrane domain traverses the lipid bilayer five times, yielding ten transmembrane helices for the homodimeric transporter

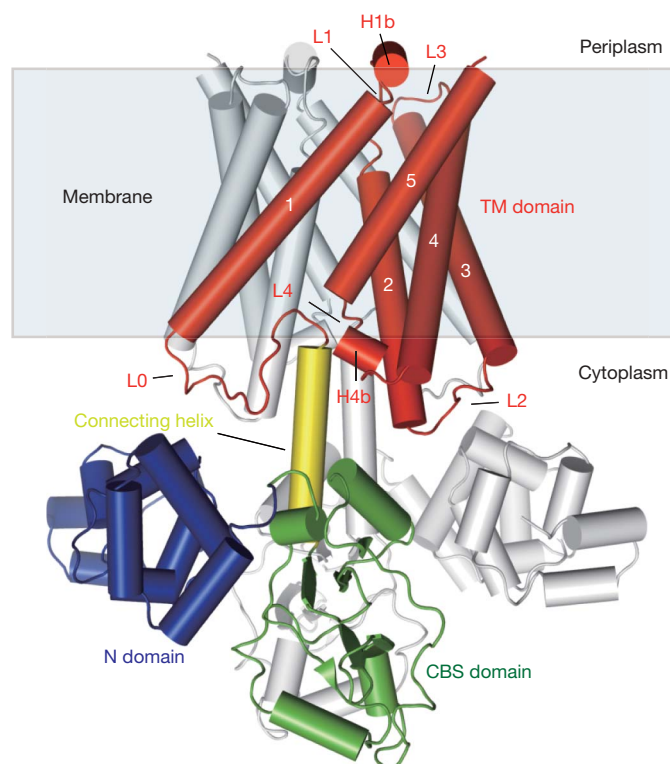


Figure 1 | Structure of the MgtE Mg^{2+} transporter. The MgtE dimer is viewed in the plane of the membrane, highlighting the N domain (blue), CBS domain (green), connecting helix (yellow), and transmembrane (TM) domain (red) in one subunit. The other subunit is coloured grey. The transmembrane helices of one subunit are numbered. The membrane surface is indicated.

¹Department of Biological Information, Graduate School of Bioscience and Biotechnology, ²Center for Biological Resources and Informatics, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama-shi, Kanagawa 226-8501, Japan. ³SORST, JST, Honcho, Kawaguchi-shi, Saitama 332-0012, Japan.

(Fig. 1). The transmembrane helix arrangement of MgtE is unlike that of any other ion channel. The transmembrane helices are connected by 5 loops (L0 to L4), in which L1 and L4 contain short helices, H1b and H4b, respectively (Fig. 1 and Supplementary Fig. 7). The transmembrane domains dimerize mainly by hydrophobic interactions, whereas hydrogen bonds between Arg 285 and Asn 402 (the other monomer) and between Tyr 273 and Glu 359 (the other monomer) stabilize the dimer formation (Supplementary Fig. 8). Arg 285 and Glu 359 are highly conserved in almost all of the eubacterial and archaeal MgtE proteins (data not shown).

Most of the transmembrane helices are kinked at a Gly or Pro residue: Gly 296 in TM1, Gly 361 in TM3, Pro 409 in TM4 and Gly 435 in TM5 (data not shown). In the CorA structure, the first transmembrane helix is twisted along the five-fold axis and kinked at the conserved Pro 303 and Gly 312 (ref. 6). The distortion of the transmembrane helices by Pro or Gly is a general feature observed in other ion channels, such as the voltage-gated potassium channel, and may provide the flexibility that facilitates the dynamic gating process¹⁹. In the transmembrane domains, acidic residues are clustered in the loop regions. Especially, in the interface between the cytosolic and transmembrane domains acidic residues protrude from both sides (Supplementary Figs 4 and 7). Only the conserved Glu 359 and Asp 432 are located in the transmembrane helices inside the lipid bilayer.

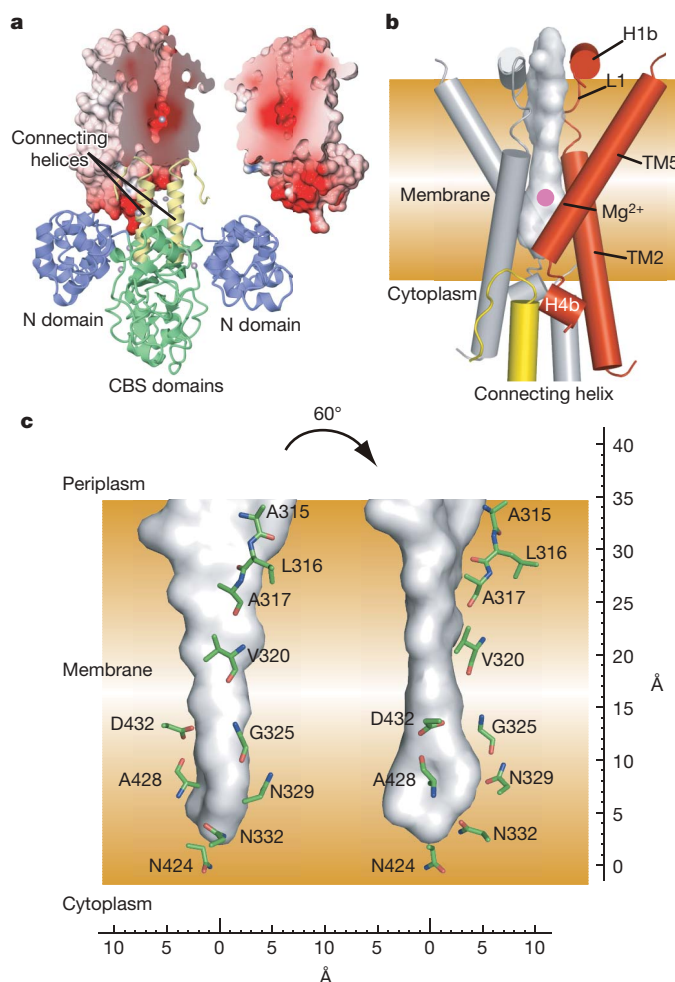


Figure 2 | MgtE pore. **a**, Cross-section view of the transmembrane domain molecular surface with an electrostatic potential (-50 (red) to 50 (blue) kT/e) and ribbon representation of the cytosolic domain. The colouring scheme is the same as in Fig. 1. **b**, Solvent-accessible surface of the pore with the pore-forming transmembrane helices. The putative Mg^{2+} is shown in purple. **c**, Residues lining the pore. The pore dimensions are also shown. For hydrophilic residues, only the side chains are shown.

The MgtE structure reveals a putative, continuous ion-conducting pathway, formed mainly by the TM2 and TM5 helices, and partly by the TM1 helices (Figs 1 and 2b), which do not traverse the membrane at the cytosolic side (Fig. 2a and Supplementary Fig. 5a). TM2, which is surrounded by the other four transmembrane helices and not exposed to the lipid bilayer, does not completely traverse the lipid bilayer, allowing the preceding hydrophobic L1 portion to adopt a non-helical, flexible structure, which forms the entrance of the pore (Figs 1 and 2b and Supplementary Fig. 9). The solvent-accessible pore is ~ 30 Å long (Fig. 2c). The diameter near the pore end at the cytosolic side is ~ 6 Å, whereas the pore width at the periplasmic entrance is broadened to ~ 15 Å in one direction (Fig. 2c). As described, the periplasmic entrance of the pore is formed by the L1 loop containing H1b (Fig. 2b and Supplementary Fig. 9) and thus a highly conserved amino acid sequence (Supplementary Fig. 7). The periplasmic side of the pore is lined with hydrophobic residues, such as Ala 315, Leu 316 and Ala 317, and the main-chain carbonyl groups of Leu 316 and Ala 317 face towards the pore (Fig. 2c). On the other hand, the cytosolic side of the pore is lined with hydrophilic residues (Asn 329 and Asn 332; Fig. 2c). These polar groups lining the pore may serve to compensate for the energetic cost of Mg^{2+} dehydration, as in other ion channels, such as KcsA and KirBac1.1 (refs 20, 21). In the middle of the pore, the side chains of the strictly conserved Asp 432 residues from both subunits face towards the centre of the pore (Fig. 3a, b and Supplementary Fig. 10a), and would presumably be critical for the Mg^{2+} transporting activity. The orientations of the side chains were assigned by the relatively clear electron density map (Supplementary Fig. 10). The pore end on the cytosolic side is apparently closed by two TM5 helices from the both subunits (Supplementary Figs 5a and 6). This closed conformation of TM5 is fixed through the aforementioned presumed interaction by shape complementarity involving TM2, H4b, TM5 and the connecting helix (Fig. 2b and Supplementary Figs 5 and 6). Therefore, the current structure seems to represent a closed state for ion conduction.

We identified five strong residual electron density peaks (4.0σ) per MgtE monomer in the full-length MgtE structure (Fig. 3). Considering that the crystallization condition included 40 mM magnesium acetate, and that all of the electron densities are close to the conserved acidic residues, it is reasonable to interpret the densities as Mg^{2+} ions, which are referred to as Mg1–5.

Mg1 is located inside the pore of the transmembrane domains, and is surrounded by the side chains of Asp 432 and the main-chain

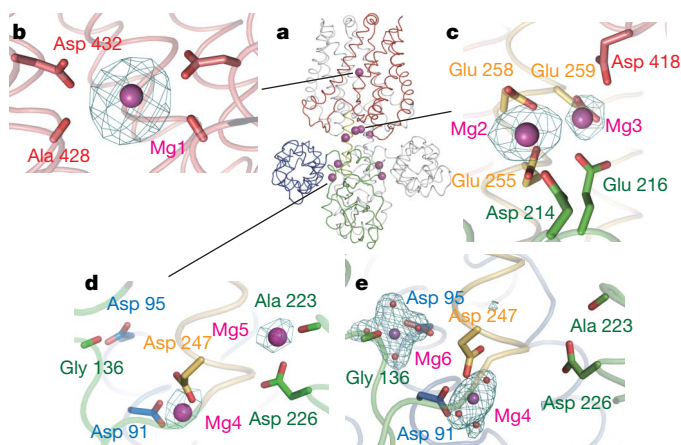


Figure 3 | Putative Mg^{2+} binding sites. The colouring scheme is the same as in Fig. 1. **a**, Side view of the overall structure with the bound Mg^{2+} . **b–d**, Closeup view of the respective Mg^{2+} (Mg1–5) binding sites in the full-length MgtE structure, with an $F_o - F_c$ simulated annealing omit map (contoured at 4.0σ) calculated with the full-length structure excluding Mg^{2+} . **e**, Close-up view of Mg4 and Mg6 in the cytosolic domain with an $F_o - F_c$ simulated annealing omit map (contoured at 4.0σ) calculated with the cytosolic domain structure excluding the Mg^{2+} and water molecules.

carbonyl groups of Ala 428 on the TM5 helices of both subunits (Fig. 3a, b and Supplementary Fig. 10a). The present Mg1 binding structure suggests that the pore functions as an 'ion-conducting pore', and may represent the selective Mg^{2+} recognition during its transport across the membrane. Asp 432 is completely conserved in the MgtE's from all three domains (Supplementary Fig. 7), suggesting that the Mg^{2+} transporting mechanism by MgtE is universally conserved. However, the present resolution of 3.5 Å does not allow an unambiguous identification of the Mg^{2+} ions bound to specific protein sites as well as a detailed discussion of the ion coordination. Further explanations of ion coordination and its selectivity are described in the Supplementary Discussion (Supplementary Fig. 11).

Mg2 and Mg3 are located at the interface between the cytosolic and transmembrane domains (Fig. 3a, c). Mg2, residing ~6 Å away from Mg3, is recognized by the side chains of Asp 214 (cystathionine- β -synthase (CBS) domain), Glu 255 and Glu 258 (the connecting helix). Mg3 is coordinated by the side chains of Glu 216 (CBS domain), Glu 259 (the connecting helix) and Asp 418 (the L4 loop of the transmembrane domain). Most of these acidic residues are conserved (Supplementary Fig. 7). In particular, the strictly conserved Asp 418 may have a key role in bridging the transmembrane and cytosolic domains (Fig. 3a, c). Both Mg2 and Mg3 would fix the connecting helix to interact with TM5 as well as with TM2 and H4b, thus locking the closed conformation of the ion-conducting pore.

Mg4 and Mg5 are located in the interface between the N and CBS domains (Fig. 3a, d). Mg4, residing ~8 Å away from Mg5, is recognized by the side chains of Asp 91 (N domain) and Asp 247 (the connecting helix). Mg5 interacts with the main-chain carbonyl groups of Ala 223 and the side chains of the Asp 226 residues in the CBS domains. Most of these acidic residues are conserved (Supplementary Fig. 7). Both Mg4 and Mg5 may not only fix the connecting helix but also stabilize the interaction between the N and CBS domains.

We further solved the structure of the cytosolic domain at 2.3 Å resolution (Fig. 4a and Supplementary Fig. 12a). The crystallographic asymmetric unit contains one molecule, which interacts with the symmetry-related molecule to form a dimer, as in the full-length MgtE structure (Fig. 4a). The dimer structure is essentially identical to that of the full-length MgtE, with an r.m.s. deviation of 0.953 Å for 492 C α atoms. In this high-resolution cytosolic domain structure, the electron density peak corresponding to Mg4 was clearly observed (Fig. 3e), with nearly hexacoordinated chemistry involving 5 oxygen atoms (two from Asp 91 and Asp 247, and three from water molecules). In contrast, the peak corresponding to Mg5 was not observed in the cytosolic domain structure, but another peak, referred to as Mg6, was observed next to Mg5 (Fig. 3e). Mg6, residing ~13 Å away from the Mg5 site, is coordinated by the side chains of the conserved Asp 95 in the N domain, the main-chain carbonyl group of Gly 136 and four water molecules, with completely hexacoordinated chemistry (Fig. 3e), which supports our interpretation of the peak as Mg^{2+} . Recently, the coordinates of the cytosolic domain of *Enterococcus faecalis* MgtE with Mg^{2+} (2.16 Å resolution) were deposited in the Protein Data Bank by a structural genomics group (accession code 2OUX; R. Sugadev, M. Sauder, S. K. Burley & S. Swaminathan, unpublished observation). The structure is consistent with our MgtE cytosolic domain structure, with an r.m.s. deviation of 2.18 Å for 233 C α atoms, although the N α 1 helices in the N domain of the *E. faecalis* MgtE are longer than those in our *T. thermophilus* MgtE. The deposited *E. faecalis* structure contains Mg^{2+} ions corresponding to Mg2, Mg4, Mg5 and Mg6 in our structure, which may substantiate our Mg2, Mg4, Mg5 and Mg6 binding sites. Mg1 and Mg3 were not observed in the *E. faecalis* cytosolic domain structure, because their recognition involves the transmembrane domain of MgtE.

We also solved the structure of the MgtE cytosolic domain in the absence of Mg^{2+} (Fig. 4a). The asymmetric unit contains two molecules (molecules A and B; Supplementary Fig. 12b). A structural comparison of the cytosolic domains in the presence and absence of Mg^{2+} revealed that the respective structures of the N and CBS

domains are essentially identical (r.m.s. deviations of 0.830 and 0.832 Å for 120 C α atoms of the N domains in the A and B molecules, respectively, and 1.00 and 1.00 Å for 104 C α atoms of the CBS domains in the A and B molecules, respectively). In contrast, the domain organization is drastically changed (Fig. 4a, b). The N domain rotates by ~120° and swings away from the CBS domain (Fig. 4a). The N domain may not have two distinct orientations depending on the Mg^{2+} concentration, but rather the cytosolic domain would be 'unlocked' and flexible in the absence of Mg^{2+} , being fixed by a crystal packing in the current structure.

Moreover, the structures of the dimeric CBS domains significantly change, with an r.m.s. deviation of 3.835 Å for 244 C α atoms (Fig. 4b). On the basis of the structural comparison of the two state cytosolic domains, we propose here the following transport mechanism. In the presence of Mg^{2+} , the CBS domains tightly dimerize and the connecting helices following are fixed by the Mg2-, Mg3- and Mg4-mediated interactions with the cytosolic and transmembrane domains, which would close the ion-conducting pore and lock the closed state (Fig. 4c and Supplementary Fig. 5). In contrast, in the absence of Mg^{2+} , the dimer interface of the CBS domains is loosened, and consequently, the connecting helices are 'unlocked' and rotated by ~20° to swing away from each other (Fig. 4b, d). It should be noted that the connecting helices are not involved in direct crystal packing interaction with symmetry-related molecules (Supplementary Fig. 12b). This movement of the connecting helices would disrupt the interactions between the connecting helices and the transmembrane domains, thus allowing the rearrangement of the pore-forming transmembrane helices and leading to the opening of the ion-conducting pore (Fig. 4d and Supplementary Fig. 5b). These interactions involve no hydrogens bond or salt bridges, except for the Mg3-mediated interaction between Asp 259 and Asp 418, which

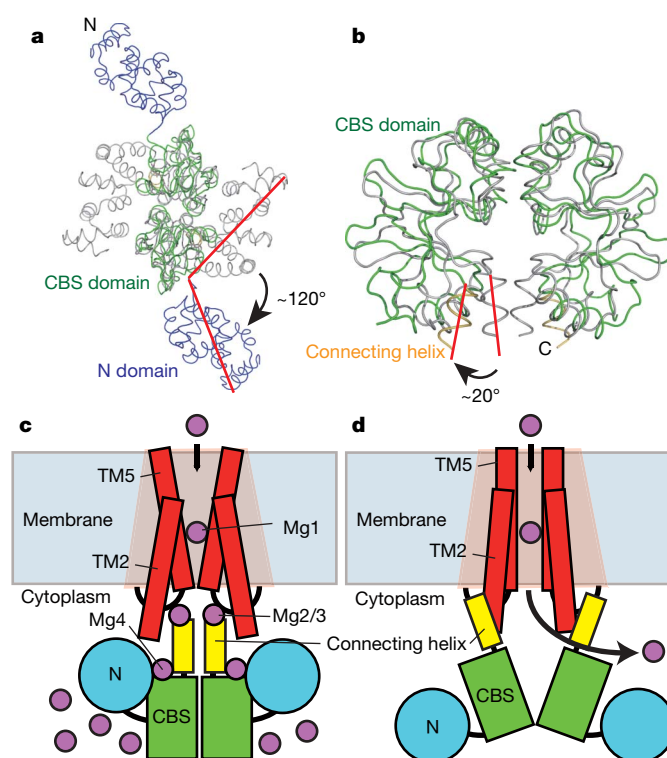


Figure 4 | Proposed Mg^{2+} homeostasis mechanism. **a, b**, Structural comparison of the Mg^{2+} -bound and Mg^{2+} -free cytosolic domains, which are superimposed on the CBS domains and viewed from the cytoplasm (**a**) and in the membrane plane (**b**). The colouring scheme of the Mg^{2+} -free cytosolic domain is the same as in Fig. 1. The Mg^{2+} -bound cytosolic domain is coloured grey. **c, d**, Proposed gating mechanism. The closed state at a high intracellular Mg^{2+} concentration (**c**). The open state at a low intracellular Mg^{2+} concentration (**d**).

would facilitate the movement of the connecting helices, after Mg^{2+} dissociation. We discuss the structural rearrangement on the periplasmic side of the pore in the Supplementary Discussion.

Altogether, the cytosolic domain of MgtE may function as a ' Mg^{2+} sensor', which regulates the gating of the Mg^{2+} transporting pore by sensing the intracellular Mg^{2+} concentration, representing a putative negative feedback or Mg^{2+} homeostasis mechanism (Fig. 4c, d). On the other hand, mammalian MgtE homologues lack the cytosolic domain (Supplementary Fig. 7), and are supposed to have a different regulatory mechanism from that of bacterial MgtE. The CorA Mg^{2+} transporter, with a fundamentally different architecture from that of MgtE, has a putative Mg^{2+} -dependent gating mechanism^{6,8}, in which the movement of the long 'stalk helix' allows the gate to open in response to changes in the intracellular Mg^{2+} concentration. It is intriguing that the two evolutionarily distinct Mg^{2+} transporters seem to have conceptually similar Mg^{2+} -dependent gating mechanisms, but with different architectures.

Thus far, the energetics of ion conductance by MgtE have not been elucidated, and no co- or counter-transported ion has been identified. The current structure seems to favour a channel-type mechanism, on the basis of the ion conducting pathway. However, the present closed conformation of the ion-conducting pore resembles the outward-facing conformation of a secondary active transporter, which suggests MgtE could act as a secondary active transporter. Further structural and functional analyses will be necessary to provide evidence to support the presently proposed model.

METHODS SUMMARY

We cloned the genes encoding full-length MgtE from seven different species into a pET28a derivative, including an N-terminal hexahistidine tag and a Precision Protease site. The MgtE transporters were overexpressed in *Escherichia coli* C41 (DE3) cells, solubilized by dodecylmaltoside, and purified by chromatography on a Ni-NTA column and a HiLoad 16/60 Superdex 200 size-exclusion column. Only *T. thermophilus* MgtE (TtMgtE) was successfully crystallized. Selenomethionine-substituted TtMgtE (10 mg ml⁻¹; 0.1% dodecylmaltoside) was crystallized by vapour-diffusion over 32–34% 2-methyl-2,4-pentanediol, 40 mM magnesium acetate, and 100 mM MES (pH 6.0)²².

The MgtE cytosolic domain (residues 1–275) was overexpressed as a fusion to an N-terminal hexahistidine tag in *E. coli* C41 (DE3) cells, and was purified by Ni-NTA column chromatography. After the His tag cleavage by thrombin, the cytosolic domain was purified by chromatography on a Mono Q column and a HiLoad 16/60 Superdex 200 column. Crystals of the cytosolic domain in the presence of Mg^{2+} were obtained by vapour diffusion over solutions containing 18–22% PEG 400, 0.2 M $MgCl_2$, and 0.1 M HEPES pH 7.4 (ref. 23). The TtMgtE cytosolic domain in the absence of Mg^{2+} was crystallized in 20% PEG 3350, 0.2 M ammonium acetate, and 0.5% octyl- β -D-glucopyranoside²³.

Diffraction data of the full-length MgtE and the cytosolic domain in the presence of Mg^{2+} were collected using an ADSC QUANTUM 315 detector on beamline BL41XU at SPring-8. Data sets of the cytosolic domain in the absence of Mg^{2+} were collected using an ADSC QUANTUM 315 detector on beamline BL5 at the Photon factory. All diffraction data were processed with DENZO/SCALEPACK²⁴ and the CCP4 suite²⁵. All structures were determined by the multiple anomalous dispersion method, using selenomethionine-substituted proteins. Details of the crystallographic procedures are given in Supplementary Information. A summary of the refinement statistics is available in Supplementary Tables 1–3.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 18 April; accepted 16 July 2007.

Published online 15 August 2007.

- Maguire, M. E. Magnesium transporters: properties, regulation and structure. *Front. Biosci.* **11**, 3149–3163 (2006).
- Nelson, D. L. & Kennedy, E. P. Magnesium transport in *Escherichia coli*. Inhibition by cobaltous ion. *J. Biol. Chem.* **246**, 3042–3049 (1971).
- Hmiel, S. P., Snavely, M. D., Miller, C. G. & Maguire, M. E. Magnesium transport in *Salmonella typhimurium*: characterization of magnesium influx and cloning of a transport gene. *J. Bacteriol.* **168**, 1444–1450 (1986).

- Bui, D. M., Gregan, J., Jarosch, E., Ragnini, A. & Schweyen, R. J. The bacterial magnesium transporter CorA can functionally substitute for its putative homologue Mrs2p in the yeast inner mitochondrial membrane. *J. Biol. Chem.* **274**, 20438–20443 (1999).
- Kehres, D. G. & Maguire, M. E. Structure, properties and regulation of magnesium transport proteins. *Biometals* **15**, 261–270 (2002).
- Lunin, V. V. *et al.* Crystal structure of the CorA Mg^{2+} transporter. *Nature* **440**, 833–837 (2006).
- Eshaghi, S. *et al.* Crystal structure of a divalent metal ion transporter CorA at 2.9 angstrom resolution. *Science* **313**, 354–357 (2006).
- Payandeh, J. & Pai, E. F. A structural basis for Mg^{2+} homeostasis and the CorA translocation cycle. *EMBO J.* **25**, 3762–3773 (2006).
- Townsend, D. E. *et al.* Cloning of the mgtE Mg^{2+} transporter from *Providencia stuartii* and the distribution of mgtE in Gram-negative and Gram-positive bacteria. *J. Bacteriol.* **177**, 5350–5354 (1995).
- Smith, R. L., Thompson, L. J. & Maguire, M. E. Cloning and characterization of MgtE, a putative new class of Mg^{2+} transporter from *Bacillus firmus* OF4. *J. Bacteriol.* **177**, 1233–1238 (1995).
- Wabakken, T., Rian, E., Kveine, M. & Aasheim, H. C. The human solute carrier SLC41A1 belongs to a novel eukaryotic subfamily with homology to prokaryotic MgtE Mg^{2+} transporters. *Biochem. Biophys. Res. Commun.* **306**, 718–724 (2003).
- Goytain, A. & Quamme, G. A. Functional characterization of human SLC41A1, a Mg^{2+} transporter with similarity to prokaryotic MgtE Mg^{2+} transporters. *Physiol. Genomics* **21**, 337–342 (2005).
- Goytain, A. & Quamme, G. A. Functional characterization of the human solute carrier, SLC41A2. *Biochem. Biophys. Res. Commun.* **330**, 701–705 (2005).
- Sahn, J., Nelson, B. & Scharenberg, A. M. SLC41A2 encodes a plasma-membrane Mg^{2+} transporter. *Biochem. J.* **401**, 505–513 (2007).
- Kowal, P., Gurtan, A. M., Stuckert, P., D'Andrea, A. D. & Ellenberger, T. Structural determinants of human FANCF protein that function in the assembly of a DNA damage signaling complex. *J. Biol. Chem.* **282**, 2047–2055 (2007).
- Ignoul, S. & Eggermont, J. CBS domains: structure, function, and pathology in human proteins. *Am. J. Physiol. Cell Physiol.* **289**, C1369–C1378 (2005).
- Bennets, B. *et al.* Cytoplasmic ATP-sensing domains regulate gating of skeletal muscle CIC-1 chloride channels. *J. Biol. Chem.* **280**, 32452–32458 (2005).
- Biemans-Oldehinkel, E., Mahmood, N. A. & Poolman, B. A sensor for intracellular ionic strength. *Proc. Natl Acad. Sci. USA* **103**, 10624–10629 (2006).
- Tieleman, D. P., Shrivastava, I. H., Ulmschneider, M. R. & Sansom, M. S. Proline-induced hinges in transmembrane helices: possible roles in ion channel gating. *Proteins* **44**, 63–72 (2001).
- Roux, B. & MacKinnon, R. The cavity and pore helices in the KcsA K^+ channel: electrostatic stabilization of monovalent cations. *Science* **285**, 100–102 (1999).
- Kuo, A. *et al.* Crystal structure of the potassium channel KirBac1.1 in the closed state. *Science* **300**, 1922–1926 (2003).
- Hattori, M., Tanaka, Y., Fukai, S., Ishitani, R. & Nureki, O. Crystallization and preliminary X-ray diffraction analysis of the full-length Mg^{2+} transporter MgtE. *Acta Crystallogr. F* **63**, 682–684 (2007).
- Tanaka, Y., Hattori, M., Fukai, S., Ishitani, R. & Nureki, O. Crystallization and preliminary X-ray diffraction analysis of the cytosolic domain of the Mg^{2+} transporter MgtE. *Acta Crystallogr. F* **63**, 678–681 (2007).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Collaborative Computational Project 4. The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the beam-line staffs at BL41XU of SPring-8 and BL5A of KEK for technical help during data collection, and Y. Sugita and T. Tsukazaki for suggestions. This work was supported by a SORST Program grant from JST (Japan Science and Technology) to O.N., by grants from MEXT to R.I., S.F. and O.N., and by grants from the Society for Research on Umami Taste, the Danone Institute, and the Yamazaki Foundation to O.N.

Author Contributions M.H. designed the research, carried out the crystallization of the full-length MgtE and the structure determinations, and wrote the paper, with editing from S.F., R.I. and O.N. Y.T. carried out the crystallization of the cytosolic domain of MgtE. S.F., R.I. and O.N. assisted with the structural determination. All authors discussed the results and commented on the manuscript. O.N. supervised the work.

Author Information The coordinates and structure factors have been deposited in the Protein Data Bank, under the accession codes 2YVX, 2YVY and 2YVZ for the full-length MgtE, and the Mg^{2+} -bound and Mg^{2+} -free cytosolic domains of MgtE, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to O.N. (nureki@bio.titech.ac.jp).

METHODS

Structure determination. First, we solved the structure of the cytosolic domain complexed with Mg^{2+} . The four selenium sites were determined with the program SnB²⁶. The selenium parameters were refined and the phases were calculated at 2.3 Å resolution using the program SHARP²⁷, with an acentric figure of merit of 0.647. The initial phases were improved by density modification using SOLOMON²⁸. The initial model was manually built using the program O (ref. 29), and interactive cycles of refinement with CNS³⁰ and manual rebuilding with O were carried out at 2.3 Å resolution. The final model contains residues 5–252, 48 water molecules and two metal ions, assigned as Mg^{2+} . The initial phases of the cytosolic domain in the absence of Mg^{2+} were calculated at 4.0 Å (an acentric figure of merit of 0.384) as the Mg^{2+} -bound cytosolic domain, and were improved by density modification using the program DM. The initial model was built using the structure of the Mg^{2+} -bound cytosolic domain, and interactive cycles of refinement with CNS and manual rebuilding were carried out at 3.9 Å resolution. The final model contains residues 5–252 from both molecules.

For the full-length TtMgtE, the 28 selenium sites were determined with the program SnB. The selenium parameters were refined and the initial phases were calculated at 4.0 Å resolution using the program SHARP, with an acentric figure of merit of 0.232. The MAD phases were applied to the peak data set and were gradually extended to 3.5 Å using four-fold averaging, solvent flattening and histogram matching in the program DM. Model building was greatly facilitated by locating the selenium atom positions and by referring to the structure of the Mg^{2+} -bound cytosolic domain. The interactive cycles of refinement with CNS, keeping a four-fold non-crystallographic symmetry restraint and manual rebuilding with O were carried out at 3.5 Å resolution. The final model contains residues 7–448 from all molecules and 18 Mg^{2+} ions. Electrostatic potentials were calculated using the program GRASP³¹. The pore in Fig. 3 was calculated by the program CAVER³². Molecular graphics were created with PyMol³³ and CueMol³⁴.

26. Weeks, C. M. & Miller, R. The design and implementation of SnB v2.0. *J. Appl. Cryst.* **32**, 120–124 (1999).
27. de La Fortelle, E., Irwin, J. & Bricogne, G. *Methods in Enzymology* (eds C. W. Carter & R. M. Sweet) 472–494 (1997).
28. Abrahams, J. P. & Leslie, A. G. Methods used in the structure determination of bovine mitochondrial F_1 ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996).
29. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
30. Brunger, A. T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
31. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).
32. Petrek, M. et al. CAVER: a new tool to explore routes from protein clefts, pockets and cavities. *BMC Bioinformatics* **7**, 316 (2006).
33. Delano, W. L. The PyMOL molecular graphics system. v.0. 97 (<http://pymol.sourceforge.net/>) (2002).
34. Ishitani, R. CueMol; Molecular visualization framework. v.1. 1 (<http://cuemol.sourceforge.jp/>) (2006).

CORRIGENDUM

doi:10.1038/nature06125

A positive-feedback-based bistable 'memory module' that governs a cell fate decision

Wen Xiong & James E. Ferrell Jr

Nature 426, 460–465 (2003)

In Box 1, equation (1) should read

$$\frac{d[A^*]}{dt} = \{\text{stimulus} \times ([A_{\text{tot}}] - [A^*])\} + f \frac{[A^*]^n}{K^n + [A^*]^n} ([A_{\text{tot}}] - [A^*]) - k_{\text{inact}}[A^*]$$

$$\text{Setting } \frac{d[A^*]}{dt} = 0, \text{ it follows that stimulus} = \frac{f \times [A^*]^n [A_{\text{tot}}] - k_{\text{inact}} K^n [A^*] - (f + k_{\text{inact}}) [A^*]^{n+1}}{([A^*] - [A_{\text{tot}}]) ([A^*]^n + K^n)}$$

which implicitly defines all of the possible steady state values of $[A^*]$ for any given value of the stimulus. The plots in Box 1 show only the stable steady states (the sections of the curves with positive slopes).

ERRATUM

doi:10.1038/nature06114

RNA-templated DNA repairFrancesca Storici, Katarzyna Bebenek, Thomas A. Kunkel,
Dmitry A. Gordenin & Michael A. Resnick*Nature* 447, 338–341 (2007)

In Figure 1, the column header indicating the repair frequencies should read “Repair frequency (Leu^+) $\times 10^{-7}$ ” rather than “Repair frequency (Leu^+) $\times 10^{-3}$ (per 10^7 viable cells)”.

CORRIGENDUM

doi:10.1038/nature06169

Structure of the *E. coli* signal recognition particle bound to a translating ribosomeChristiane Schaffitzel, Miro Oswald, Imre Berger, Takashi Ishikawa,
Jan Pieter Abrahams, Henk K. Koerten, Roman I. Koning & Nenad Ban*Nature* 444, 503–506 (2006); doi:10.1038/nature05182 (published online 29 October 2006)

During the preparation of the manuscript, we inadvertently mislabelled ribosomal protein L32 as ribosomal protein L18 when interpreting the density based on the 50S coordinates (PDB accession number 2AW4). Therefore, whenever L18 is mentioned in the text and in Figs 3 and 4, it should be considered to refer to ribosomal protein L32. Our results and conclusions are not affected.

naturejobs

JOBS OF THE WEEK

Since the rise of biotechnology in the 1970s, the worlds of bioscience and business have drawn ever closer together. But for jobseekers interested in the biosciences industry, how much of an asset is business training? It may offer some advantages. The pharmaceutical industry, for example, is suffering serious job losses (see *Nature* **448**, 965; 2007), but there still are jobs available, and business training is one way to get a leg up. The extent to which that training translates into broadened opportunities is hard to quantify. The Keck Graduate Institute in Claremont, California, which specializes in bioscience degrees that incorporate business skills, has released a report that assesses the situation (see www.kgi.edu/x6503.xml).

Molly Schmid and Helen Liu looked at the 3,790 jobs that were open in July at the top five revenue-generating drug companies (Pfizer, GlaxoSmithKline, Sanofi-Aventis, Merck and Johnson & Johnson) and the top four biotechnology firms (Amgen, Genentech, Biogen IDEC and Genzyme). The results may surprise life-science researchers.

Roughly 17% of the jobs were for positions in research and development (R&D); the rest were for jobs related to the business side of the companies. Of these, 7% were in manufacturing, 24% were in administration (including duties such as information technology, finance and legal services), 25% were regulatory jobs (such as quality control, technical writing and statistical analysis), and 27% were in sales and marketing.

Schmid says the results are a bit of a wake-up call. "Academics don't typically see or think about what types of jobs are available in industry," she says. "Many will be truly surprised. They probably thought there'd be more R&D."

The study is, of course, just a snapshot of one month. But the numbers underscore the importance of having multiple skill sets. "Students need to appreciate the types of jobs for which their science educations might be preparing them," says Schmid. "And they might not be getting that message in their organic chemistry labs."

Gene Russo, acting editor of *Naturejobs*

CONTACTS

Acting Editor: Gene Russo

European Head Office, London

The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:

Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com

Natureevents:

Claudia Paulsen Young
(+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com

France/Switzerland/Belgium:

Muriel Lestringuez (4994)

Southwest UK/RoW:

Nils Moeller (4953)

Scandinavia/Spain/Portugal/Italy:

Evelina Rubio-Hakansson (4973)

Northeast UK/Ireland:

Matthew Ward (+44 (0) 20 7014 4059)

North Germany/The Netherlands:

Reya Silao (4970)

South Germany/Austria:

Hildi Rowland (+44 (0) 20 7014 4084)

Advertising Production Manager:

Stephen Russell
To send materials use London
address above.

Tel: +44 (0) 20 7843 4816

Fax: +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Jasmine Myer

US Head Office, New York

75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo

Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:

Ayako Watanabe
Tel: +81-3-3267-8765
e-mail: a.watanabe@natureasia.com

**Professor in
Physiological Chemistry**
Johannes-Gutenberg
University Mainz
Mainz (Germany)
Turn to page 7

Faculty Position
Harvard University -
FAS Center for
Systems Biology
Boston, MA (USA)
Turn to page IFC

W3 Professor
Immunology
University of
Hohenheim - Institute
of Nutritional Medicine
Stuttgart (Germany)
Turn to page 15

**Postdoctoral
Fellowships**
MRC - Epidemiology
Unit
Cambridge (UK)
Turn to page 1078

Faculty Position
MIT - Department
of Biology
Cambridge, MA (USA)
Turn to page 11

A new note for Nat

Rock on.

Gareth Owens

Nat used to be in a band. Until yesterday that was. That's him, sat in the armchair, the unhappy looking one talking to Midge. And there's Duffy and Slingshot. And that's me in the other corner.

You can't always see me on the screen, Tim was holding the camera. I'm only the manager, and it was the band we wanted to catch. This was a promo, backstage thing for the website.

This was taken about three years ago, back in the day when Invisible Ear was the biggest band on three continents. You must have been living under a rock to have missed them, even if you couldn't actually hear anything they did. They were the crest of a wave, first of a new breed, pioneers on the new frontier. And I can honestly say that I've never heard a single note that they've ever played. Not that I'm that bothered, they have made me a very happy man over the past few years. M. Osborne, manager to the stars, known to one and all as Mozzy. I have bands queuing up for me now, and I owe it all to Nat and the gang.

But there would have been no Invisible Ear at all, no career for Nat, or for me, if it hadn't been for an irritated shopkeeper.

The problem this guy had was he wanted to discourage gangs of youngsters from hanging around his shop, putting off his paying customers. His solution was to put a buzzing box by the door that made an irritating high-pitched sound that only the youngsters could hear. 'Dults lose their hearing from the top down, so the older customers could come in and spend money without any problem, but anyone under the age of about 18 would have this constant buzz, like toothache. Really put them off.

Well, kids ain't stupid. They took to this idea that there's stuff they can hear that the 'dults can't and made it their own. Suddenly there were ring-tones for cell phones just for them, and that was when Nat and Slingshot had the idea to get together and begin Invisible Ear. They started writing tunes in this register. A music that was theirs and theirs alone. No grown-ups allowed, 'dults keep-out.

I was always on the lookout for the next

big thing, and I saw Invisible Ear play at a school hall. I wasn't expecting much. School bands are usually awful, shouting boys doing covers of angry music, but this was something else. Nat raised his hands in the air, counted the band in with a one-two-three and they started to play. The audience went wild. I could see heads bobbing up and down, and they danced around to this invisible beat, and yet I could hear nothing, nada, zip and zilch.



I stood amazed and puzzled. For the next three minutes the band played, and Nat sang (he recorded vocals prior, pitch shifted them up and then mimed so it looked like he was singing). Then they stopped. The hall, which had been silent while the band played, erupted into wild shouts of unrestrained enthusiasm.

I signed them like a shot. For the first time, here was a band I could represent that I didn't actually have to listen to.

But the real fun was taking the demo around to record companies, playing these ultrasonic songs to the guys in suits, and then telling them that only cool people could hear the music. Hans Christian Anderson would have loved it.

Screwing their faces up with concentration they'd lean forward and then they

would nod, and look like they were getting into it, and about half the time Nat hadn't switched the player on.

Invisible Ear took the world by storm. Not only music but vids, games and online worlds, Japanese children would queue for hours just to get the latest Ear game to play at home. They became a worldwide cultural phenomenon.

Then yesterday, at the height of the band's fame and popularity, they sacked Nat. He's too old you see. He sat in that chair there and the poor kid was almost in tears. He's been wearing hearing aids for the past six months. He was totally distraught, been tryin' to hold off the inevitable, but what can you do? He's over the hill four months before his eighteenth birthday. I told him: "Nobody can keep time at bay, not even you Nat. You founded a club meant to keep grown-ups out — and then you grew up. Sorry."

I let him wallow for a while, but to be honest what kind of manager would I be if I hadn't seen this coming?

"Look, Nat. Take those things off your ears, you are not going deaf." I didn't give him the choice. I walked over to where he sat and pulled the hearing aids off his ears, he yelped a bit. "Like I said, you founded a club for the youth market, but you're not the only one getting older. The record company is beginning to get complaints that the recordings they made of Invisible Ear's songs have faded. Your fans are putting on your music, and they can't hear it any more. It's not the songs that are fading — your first lot of fans are about the same age as you, so they can press the play button as much as they like but nothing's gonna come out, at least not for them anymore."

"So?" he said.

"It's time for Nat to grow up and go solo," I said grinning. "To transpose all your songs into an audible register, re-record them as a series of duets with the hottest stars, and get your fans to buy them all over again. Welcome to the 'dults Nat, I think you'll find it much more fun than you were expecting."

Gareth Owens is a palaeolinguist and a student of the arcane writings of ancient and lost civilizations. He is a composer, writer and traveller and has recently completed a novel.

JACEY